

Reading Recovery: A Sector-Wide Analysis

Centre for Education Statistics and Evaluation



Publication and contact details

Centre for Education Statistics and Evaluation 2015,
Reading Recovery: A Sector-Wide Analysis
Report prepared by D Bradford and W Wan. This report is available at:
<http://www.cese.nsw.gov.au/publications/reports-and-presentations>

Authors

Deborah Bradford
Principal Evaluator, Evaluation and Major Projects
Centre for Education Statistics and Evaluation
T 02 9561 8661

Wai-Yin Wan
Senior Data Analyst, Evaluation and Major Projects
Centre for Education Statistics and Evaluation
T 02 9561 8871

The authors would like to acknowledge the contribution of staff in the Centre for Education Statistics and Evaluation, who contributed to the design and conduct of this research and the publication process. We would also like to acknowledge staff in the Department's Early Learning and Primary Education Directorate, who provided data and gave feedback on the research. Finally, we would like to thank the independent peer reviewers for their helpful feedback.

For more information about this report, please contact:
Centre for Education Statistics and Evaluation
Department of Education
GPO Box 33
SYDNEY NSW 2001
Email: cese@det.nsw.edu.au
Telephone: +61 2 9561 1211
Web: www.cese.nsw.gov.au

Contents

1. Background	2
2. Method	5
3. Results	13
4. Discussion	21
References	25
Appendix A. Sample Representativeness	27
Appendix B. Year 1 Regression Technical Information	31
Appendix C. Propensity Score Matching	34
Appendix D. Year 3 Regression Technical Information	36
Appendix E. Year 1 Supplementary Results	37
Appendix F. Year 3 Supplementary Results	42

1. Background

The New South Wales (NSW) Government has a strong focus on improving students' literacy and numeracy skills in the early years of schooling. In 2011 the NSW Government committed \$261 million under the *Literacy and Numeracy Action Plan* to improve literacy and numeracy for students in the early years (Kindergarten to Year 2). This focus on early literacy is important as students who are not reading well by the time they reach Year 3 face significant challenges for the remainder of their schooling (Willms 2003). Early identification of students who are having difficulty reading and the introduction of effective remediation strategies are both critical elements of a school's role in developing the reading capabilities of their students.

Reading Recovery Overview

There are many different interventions used in NSW primary schools to assist young students improve their literacy outcomes¹. One of the most widely used interventions is Reading Recovery (RR), which has been at the forefront of the remediation effort in the NSW Department of Education for over two decades. In 2012, RR was offered in over half (approximately 60%) of NSW government primary schools, with approximately 14% of all Year 1 students participating in the intervention.

RR was developed in New Zealand in the 1970s by Dame Marie Clay as an intensive individualised literacy intervention that aims to accelerate literacy learning for students performing in the bottom 20 per cent of Year 1 (Department of Education and Communities 2015; What Works Clearinghouse 2008). RR tuition is provided on a one-to-one basis over 12-20 weeks with the intention of raising students' performance to the average level of their Year 1 peers, thereby enabling them to benefit from classroom instruction and sustain achievement throughout the early years of school (Department of Education and Communities 2015; May et al. 2013; 2015).

While the intention of RR is to be responsive to each individual student's needs, the intervention typically addresses several aspects of reading and writing processes that support the comprehension of texts. These include: vocabulary, fluency, comprehension, writing, phonemic awareness, phonics, motivation and oral language (What Works Clearinghouse 2013)². RR does not claim to align itself explicitly with a particular classroom-based approach and is substantively based on the notion that students draw on multiple sources of information (e.g., visual, linguistic, text-based) when learning to read (Chapman & Tunmer 2011; Reading Recovery Council of North America 2015, see <http://readingrecovery.org/>).

Eligibility for RR is assessed using the Observation Survey of Early Literacy Achievement, a standardised assessment based on teacher observations of student performance on six tasks related to early literacy skills (Clay 2002; Reading Recovery Council of North America 2015, see <http://readingrecovery.org/>). These include: Letter Identification, Word Test, Concepts About Print, Writing Vocabulary, Hearing and Recording Sounds in Words, and Text Reading (see <http://readingrecovery.org/reading-recovery/teaching-children/observation-survey>). The aim of this teacher-administered assessment is to identify the lowest 20 per cent of text readers in Year 1.

Following the administration of the Observation Survey, final selection of participating students is made in consultation with the school. Students begin participating in RR continuously across a school year when a place becomes available with a trained RR teacher. They 'successfully discontinue' RR when they have achieved the average reading level for their grade, typically a score of 16 or above on a re-test of the Observation Survey, as administered by an independent assessor. Students who do not reach this level after 12-20 weeks are referred for further specialist support or for long-term literacy support.

1 Some of the other most commonly used interventions in NSW include MiniLit; QuickSmart Literacy; Accelerated Literacy; First Steps; Language, Learning and Literacy; Literacy on Track, Literacy Lessons; Focus on Reading, Off to a Good Start: Learning to Read K-2 (OTAGS); Principals as Literacy Leaders (PALL); Reading Matters; Reading to Learn; and MultiLit (Australian Council for Educational Research (ACER) 2013).

2 Further information on the content of RR lessons in NSW government schools can be found here: http://www.curriculumsupport.education.nsw.gov.au/earlyyears/reading_recovery/lesson.htm.

Students who do not complete their series of lessons within a calendar year may have their lessons carried over to the next year. Students may also stop participating in RR if they transfer schools and are not able to continue with RR lessons at their new school.

Evidence for the Effectiveness of RR

A considerable amount of research has been conducted world-wide examining the impact of RR on student outcomes. Unfortunately, most of the studies assessing the effectiveness of RR have employed relatively weak research methodologies, which calls into question the reliability of the evidence (D'Agostino & Murphy 2004; May et al. 2015; What Works Clearinghouse 2008; 2013). Among 202 studies reviewed by the What Works Clearinghouse (2013), only three studies, all Randomised Controlled Trials (RCTs), met the highest evidence standards (i.e. were considered strong sources of evidence). The evidence derived from these studies suggests that RR is an effective intervention for increasing student literacy levels. In particular, results from this review revealed positive RR effects on general reading achievement and potentially positive effects on alphabets, reading fluency and comprehension (What Works Clearinghouse 2013).

More recently, the effects of RR on student outcomes were examined in a multi-site RCT conducted as part of a large-scale initiative to expand RR to more high-need schools throughout the United States (May et al. 2015). This study included 433 matched pairs of students (drawn from 147 schools) that were identified as eligible for RR. Within each pair, one student was randomly allocated to the treatment group (RR + classroom instruction) while the other was assigned to the control condition (classroom instruction + an optional non-RR intervention). Student outcomes were examined using the Iowa Test of Basic Skills, completed following the 12-20 week RR intervention period, whilst controlling for the students' pre-test performance. Results showed a significant positive effect of RR on overall reading scores, with an average effect size of 0.69 for RR relative to control students who were eligible for RR but did not participate in the program (May et al. 2015).

These positive outcomes are also consistent with the conclusions outlined in a recent report commissioned by the NSW Ministerial Advisory Group on Literacy and Numeracy. The report concluded that evidence was lacking in relation to most of the literacy and numeracy interventions included in the review. RR was one of the few interventions with a reasonably strong evidence base for its efficacy (ACER 2013). This is not to suggest that other approaches to literacy and numeracy instruction or remediation are ineffective, just that there is little evidence upon which to compare the effectiveness of the different approaches.

While the evidence reviewed above provides reason to be cautiously optimistic about the effectiveness of RR, these studies are not without their limitations. Each of the studies included in the What Works Clearinghouse review had relatively small sample sizes (fewer than 100 students). Program effects are not necessarily generalisable from small, non-representative samples to whole school populations. Although May et al. (2015) observed positive RR effects among a much larger cohort, sample attrition (31%) was relatively high, raising the possibility that the students lost to follow-up differed in critical ways from those who remained in the study. This may have led to biased estimates of the system-wide effectiveness of the program. Furthermore, even though RCTs are the best way of estimating true and unbiased treatment effects, one of their limitations is that the control groups either receive no supplemental intervention or a constrained alternative. Both of these alternatives are unlikely to represent how educators respond to low-performing students in the absence of RR.

In addition, the outcomes from some RR evaluation studies have been questioned on the basis of the measures used to assess program effectiveness (Reynolds & Wheldall 2007). Among the studies included in the What Works Clearinghouse (2013) review, literacy outcomes were assessed at least in part using sub-tests from Clay's Observation Survey. Some researchers have argued that this instrument is not a sufficiently independent measure of literacy progress and that the sub-tests align very closely with the content delivered within RR lessons (Grossen, Coulter & Ruggles 1997; Reynolds & Wheldall 2007). As a result, this measure could potentially inflate learning effects for RR students, who may continue to be challenged by novel reading materials.

RR has also been subject to extensive criticism, particularly in New Zealand, where it has formed a key part of the national literacy strategy for over 25 years (Chapman & Tunmer 2011; Greaney 2011; Tunmer et al. 2013). Tunmer et al. (2013) point out that RR has failed to lift the literacy skills of the poorest

performing students, evidenced by the fact that there has been no closing of the achievement gap between low and high performing readers in the 25 years the program has been operating. In support of their argument, Tunmer et al. observe that students at higher starting points in reading achievement are most likely to be successfully discontinued from the program, and that gains among these higher performing students are not sustained (Chapman & Tunmer 2011; Moats 2007; Tunmer & Chapman 2003; Tunmer et al. 2013). While these are important considerations, it is perhaps no surprise that students at higher starting points are more likely to complete the program; they simply require less improvement before they are successfully discontinued. The effectiveness of RR for students at the lowest starting points can only be ascertained by comparing growth among students at similar starting points who do not receive RR. The sustainability of any benefits associated with RR, on the other hand, is a critically important issue that has not been rigorously examined to date.

Objectives of the Current Study

While the balance of the evidence suggests that RR is an effective intervention for raising student literacy levels, most evaluations of RR have been conducted outside Australia. This raises the question of whether RR is equally effective in NSW. Programs can be implemented in different ways across jurisdictions, which can lead to variability in the outcomes achieved. Moreover, the capabilities of teachers and the other strategies employed to remediate low levels of literacy may differ greatly across countries. Second, the extant evidence has not resolved the issues raised by critics regarding the effectiveness of RR for low performing students. Further research that carefully accounts for student baseline achievement is needed to assess whether RR is differentially effective for students at low versus high starting points. Finally, the long-term sustainability of the results achieved by RR have not been well considered using rigorous methodologies in the existing literature.

In light of these concerns, a rigorous and up-to-date sector-level evaluation of RR is critically important. This is particularly important considering the current educational policy environment in NSW. Under *Local Schools, Local Decisions*³, school leaders have much greater authority to make local decisions about the programs that best suit the needs of their schools. While the evidence suggests that RR is an effective intervention in some contexts (e.g. compared to no supplemental intervention), understanding more about how RR works, for whom and under what circumstances will provide principals with the information they need to make informed decisions at the local level.

The primary objectives of the current evaluation were to determine the impact of RR on students' literacy outcomes and whether any benefits associated with participating in RR are sustained over the longer term. This analysis was conducted at the sector-level (across NSW government schools) and focussed on identifying the impact of RR compared to students who had similar characteristics but who attended a school that did not offer RR. An important aim of the current study was to determine whether there was any interaction between baseline achievement levels and the effectiveness of RR. The key research questions addressed in this evaluation were:

1. What proportion of students participating in RR reach the minimum reading levels expected of Year 1 students, and achieve literacy outcomes equal to or greater than those of their peers?
2. In the short-term, are literacy outcomes for students who participate in RR greater than those for comparable students who do not participate in RR?
3. Are any benefits of RR sustained over longer periods of time (i.e. to Year 3)?

3 See: <http://www.schools.nsw.edu.au/media/downloads/news/announcements/yr2011/aug/local-schools-local-decisions.pdf>.

2. Method

Design

The study employed a quasi-experimental design drawing on retrospective data that detailed participation in RR and student outcomes in the early years of school. Outcomes for students who participated in RR (treatment group) were compared to a group of students with similar starting literacy levels and socio-demographic characteristics who did not participate in RR (control group)⁴.

Student outcomes were measured at two time-points to assess the impact of RR over the short- and longer-term. All NSW government school students are assessed on their literacy and numeracy capabilities at entry to Kindergarten by way of the Best Start assessment. However, there is currently a lack of mandatory formal assessment between the Best Start assessment and the commencement of NAPLAN in early Year 3. This complicates analysis of the short-term impact of RR at the end of Year 1, as there are no state-wide benchmarks with which to compare performance of students who receive RR versus those who do not. Some schools voluntarily track student progression between Best Start and NAPLAN using a curriculum-referenced Literacy Continuum. Data from the Literacy Continuum were used in the current study to assess literacy progression to the end of Year 1. The longer-term effect of RR was examined using the NAPLAN Reading assessment conducted in Year 3. These measures are described in more detail below. All student outcomes were analysed at the individual level.

Data Sources

Three sources of student-level data were used in this evaluation.

Reading Recovery Data

RR teachers collect a wide range of data for each student participating in the intervention. This includes student- and school-level demographic information, duration of program participation (i.e. number of weeks and lessons), RR program status (discontinued, referred, transferred, and carried-over), and student performance on three assessment tasks: Clay's Observation Survey, the Burt Word Reading Test and a Writing Vocabulary task. Each assessment is administered to students prior to program entry and again following delivery of RR. For the current study, data for the cohort of students who undertook RR in 2012 was used to compare achievement outcomes to a cohort of students who did not undertake RR in 2012 (as described in further detail under the section *RR and Non-RR Cohort Selection*).

Students were included in the 2012 RR year dataset if they had a non-missing student ID and had an initial score on Clay's Observation Survey plus a subsequent Survey assessment score in the identified calendar year. This included students who successfully discontinued from RR, those who were referred for further specialist support and/or long-term literacy support, as well as those transferred out of the program or who had their lessons carried over to the following year. Students who were transferred or carried-over into another school year (approximately 5% of RR students in each year) were excluded from the sample because there were too few students in these groups to estimate differential treatment effects. All analyses of RR outcomes focused on students who were deemed to have completed the program and exited with either a status of 'successfully discontinued' or 'referred' for long-term literacy support.

4 Students in the control group were drawn from schools that did not offer RR in 2012.

Best Start and Literacy Continuum K-10 Data

The Literacy Continuum K-10 was developed by the NSW Department of Education as a professional learning and teaching tool that can be administered by teachers to help describe the development of literacy skills among students. The Continuum contains eight evidence-based aspects of literacy learning that are regarded as critical to literacy success: Reading Texts, Comprehension, Vocabulary Knowledge, Aspects of Writing, Aspects of Speaking, Phonics, Phonemic Awareness and Concepts about Print (for more detail on the aspects see: <http://www.curriculumsupport.education.nsw.gov.au/literacy/>). Each aspect is organised into a series of sequential clusters that describe the range of skills expected of students across the years of schooling. Within each cluster, a set of markers are described that reflect the milestones of expected achievement at any particular point in time.

As student literacy skills develop within each aspect, they are expected to move progressively across clusters. From Kindergarten to Year 6, the clusters within each aspect range from 1 to 12, with the exception of Phonics (8 clusters), Phonemic Awareness (6 clusters) and Concepts about Print (5 clusters). However, the clusters are not distributed equally across school years. Literacy skills expected in Kindergarten are encompassed in the first four clusters, representing the rapid development of literacy skills typically achieved in that time. Expected achievement for Year 1 and Year 2 students is captured in two clusters, while subsequent years are constrained to one cluster per year⁵.

Assessment on the Continuum at the beginning of Kindergarten (i.e. Best Start assessment) has been mandatory in all government primary schools since 2010. This assessment aims to help teachers identify the skills that each student brings to school and includes 7 of the 8 Continuum aspects (Vocabulary Knowledge is excluded). Following the Best Start assessment, use of the Continuum is not mandatory and is used at the discretion of schools and teachers.

In the current analysis, Literacy Continuum data were sourced for all students who had been tracked against the Continuum at three time points:

- Best Start Literacy assessment data collected at Term 1, 2011 was used to help identify a cohort of students who did not participate in RR in 2012. These data were also used as a baseline indicator of student performance and as a source of student demographic information⁶.
- Literacy Continuum data collected at the end of Term 4 of Kindergarten (T4K) in 2011 was used as a source of student demographic information and as a baseline indicator of student performance prior to the availability of RR in Year 1.
- Literacy Continuum data collected at the end of Term 4 of Year 1 (T4 Y1) in 2012 was used as a source of student demographic information and to examine outcomes for RR and non-RR students at the end of Year 1.

NAPLAN Data

To examine the longer-term effects of participating in RR, performance in the 2014 NAPLAN Year 3 Reading assessment was analysed for RR and non-RR students. The primary aim of this assessment is to measure literacy proficiency through the interpretation of written text (see: <http://www.nap.edu.au/naplan/reading/reading.html>). The current analysis focussed on scaled student scores, which ranged from -20.4 to 771.2 in 2014^{7,8}. More information on the NAPLAN assessment is available at: <http://www.nap.edu.au/naplan/naplan.html>.

5 See: http://www.curriculumsupport.education.nsw.gov.au/literacy/assets/pdf/continuum/K-10_lit_cont_overview.pdf.

6 Student demographic information was sourced from the Best Start assessment and from the Literacy Continuum assessments conducted at the end of Kindergarten and the end of Year 1. Where there were disagreements in information at each time point, the value that was recorded most often was selected.

7 NAPLAN score equivalence tables for 2014 can be found at: http://www.nap.edu.au/verve/_resources/NAPLAN2014_EquivalenceTables.pdf. These tables show the range of NAPLAN raw scores, scaled scores and associated performance bands for each domain across Years 3, 5, 7 & 9. Table 2 shows the scaling for the Year 3 Reading domain.

8 The process of transforming raw NAPLAN scores to scaled scores can yield negative values. NAPLAN scales are 'constructed so that any given score represents the same level of achievement over time' (see: <http://www.nap.edu.au/results-and-reports/how-to-interpret/score-equivalence-tables.html>).

RR and Non-RR Cohort Selection

To determine the impact of RR on student performance, outcomes for students participating in RR and a control group of non-RR students were compared at two time-points: at the end of Year 1 on Literacy Continuum measures and in Year 3 on NAPLAN Reading scores. For these analyses, the cohort of students who participated in RR in 2012 ($n=9,587$)⁹ was selected as the RR sample. This cohort was selected because sufficient data were available to identify a potential control group using the 2011 Best Start assessment, and to measure Literacy Continuum outcomes at the end of Year 1 (2012) and NAPLAN outcomes in Year 3 (2014).

In the 2011 Kindergarten Best Start assessment, records were available for 65,535 students, of whom 8,638 participated in RR in 2012 and 56,897 did not. These records were then matched with Literacy Continuum records collected at the end of Term 4 in Kindergarten to gather important indicators of baseline literacy progress prior to the availability of RR in Year 1¹⁰. Matching to the Kindergarten Term 4 Literacy Continuum records resulted in a loss of 2,837 student records from the overall Best Start 2011 sample, leaving 8,396 RR and 54,302 non-RR student records.

To measure short-term outcomes of RR on student performance, the dataset was merged with Literacy Continuum records collected at the end of Year 1. This reduced the sample to 8,271 RR students and 50,290 non-RR students. To measure the relative long-term sustainability of the impact of RR, the dataset was then merged with NAPLAN Year 3 Reading data collected in 2014, which reduced the sample to 7,573 RR students and 46,841 non-RR students. Next, the RR sample was restricted to include only those students who completed RR in 2012 and had a status of 'discontinued' or 'referred' at program exit. This reduced the RR sample to 7,208 students. Finally, to minimise the risk of selection bias, the non-RR sample was reduced by selecting only those non-RR students who attended non-RR schools, leaving 13,321 non-RR students.

It is important to note that these samples reflect matched data records, and there were many cases where student records were matched across datasets, but the data for the required measures were missing or unknown. This was particularly the case for the Literacy Continuum data collected at the end of Kindergarten and at the end of Year 1, neither of which are mandatory for schools to administer. This resulted in the loss of a number of cases in analyses of student outcomes at the end of Year 1 on Literacy Continuum measures and at the end of Year 3 on the NAPLAN Reading assessment. Notably, this attrition was not equal across RR and non-RR cohorts, as non-RR students were less likely to have complete Literacy Continuum records from Best Start through the end of Year 1.

While this may be considered a potential source of bias, the assumption made in using Continuum data as a key outcome variable is that the factors that lead to the use of the Continuum at each time point are the same across RR and non-RR schools. These factors could include employment of school and/or instructional leaders who have a focus on using data to differentiate teaching, or participation in other programs where use of the Continuum is mandated (e.g. The Department of Education's *Early Action for Success Strategy*). There was no a priori reason to assume that these factors led to systematic biases in the comparison of student outcomes in RR and non-RR schools.

The process of sample selection is shown in Figure 1. This presents the RR and non-RR cohorts from those first identified in the 2011 Best Start assessment to the final numbers included in the analyses of student outcomes at the end of Year 1 and in early Year 3. Student outcomes at the end of Year 1 were examined separately for each Literacy Continuum aspect (as described in further detail under the section *Modelling the Impact of RR on Year 1 Outcomes*). As an example, Figure 1 presents the number of students included in the analysis of the aspect corresponding to Reading Texts. While there were small differences in the number of students included in each of the remaining Literacy Continuum aspect analyses, a very similar selection pattern was observed across all aspects.

Appendix A provides information bearing on the extent to which the final samples included in statistical analyses were representative of the full cohort of students participating in the 2011 Best Start assessment. While there were slight variations in the characteristics of students included in the analyses compared to the full 2011 Best Start cohort, the samples were deemed sufficiently similar to be confident in the robustness of the analysis.

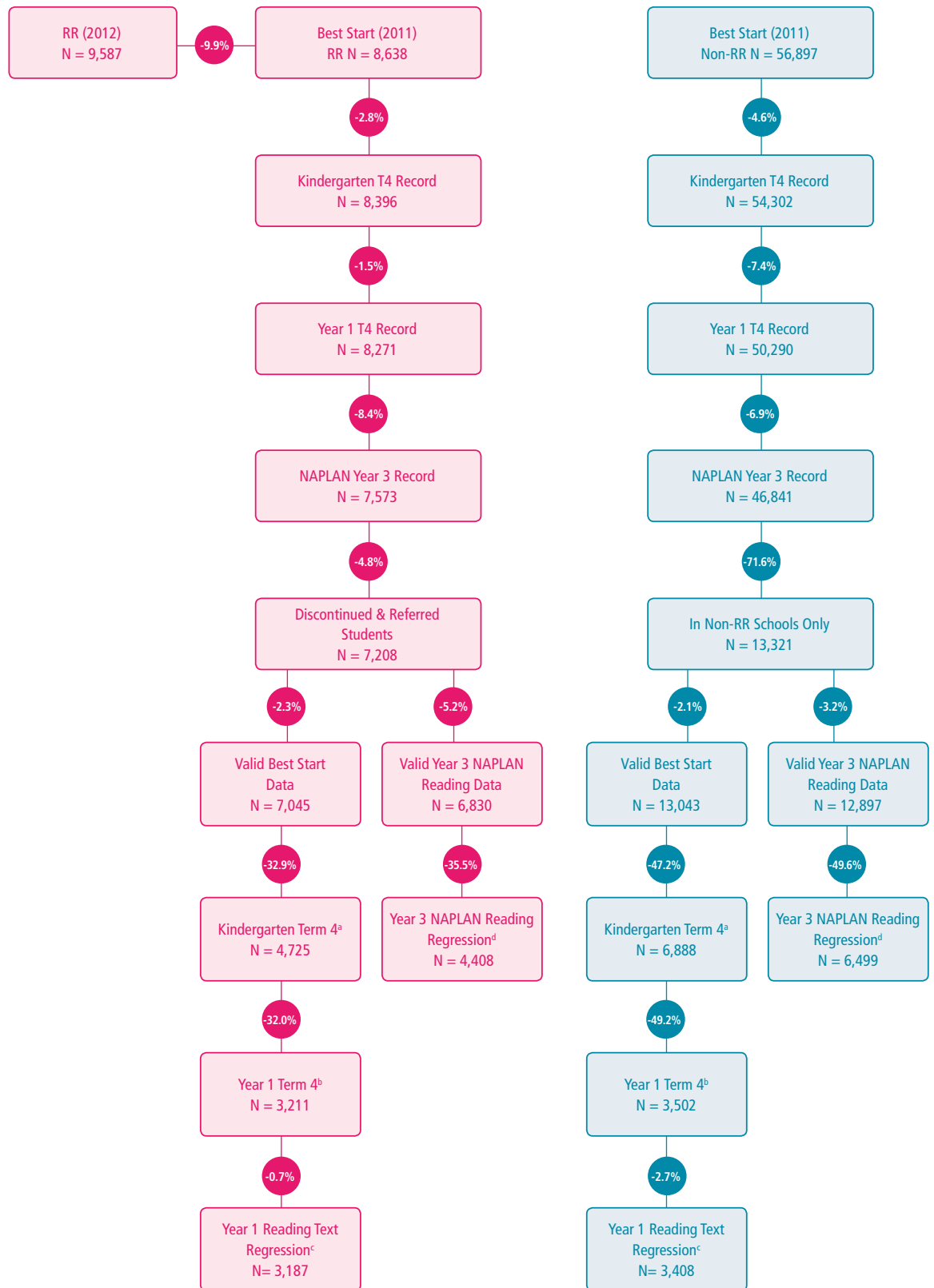
9 This cohort included students with non-missing Student Reference Numbers (SRNs). This included 164 students who started RR in 2011 and had their lessons carried over to 2012, and 9,423 students who started RR in 2012.

10 All data matching was done on the basis of student identification records, with student IDs (SRNs) used as the primary matching criterion. Names and dates of birth were used, as required, as secondary matching criteria.

Figure 1:

Selection of RR and non-RR cohorts

Sources: NAPLAN data extracted from the Statistics Unit, Centre for Education Statistics and Evaluation (CESE). Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, Department of Education (DOE).



- ^a Includes students with valid (i.e. non-missing scores) on the Reading Text aspect of the Literacy Continuum at Best Start and at Term 4, Kindergarten.
- ^b Includes students with valid scores on the Reading Text aspect of the Literacy Continuum at Best Start, Term 4, Kindergarten and Term 4 Year 1.
- ^c The Year 1 Reading Text regression includes students with valid Reading Text scores at each Continuum time point, excluding those cases where the difference between scores at the end of Kindergarten and the end of Year 1 was negative (by 2 levels or greater) or where the socio-demographic information used in the regression was missing.
- ^d The Year 3 NAPLAN regression includes those with valid Reading Text and Comprehension scores at Best Start and Term 4, Kindergarten, excluding those where the socio-demographic information used in the regression was missing.

Note. The number of valid records at each time point varies slightly for each Continuum aspect.

Comparisons of Final RR and Non-RR Cohorts

Table 1 shows the demographic distribution for the final groups of RR and non-RR students who were included in the regression analyses of student outcomes at the end of Year 1 and at NAPLAN in Year 3¹¹. There were slight differences in the demographic characteristics of RR and non-RR cohorts across both regression samples. Compared to non-RR students, RR students in both samples were significantly more likely to be male, Aboriginal, to have a non-English speaking background (NESB), and to be classified as English as a Second Language (ESL) Phase 1 (all p-values < .001). With regards to parental occupation and education level, students who did not participate in RR were significantly more likely to have parents who completed Year 12 and were employed in management positions relative to RR students (p-values < .001).

Table 1:
Demographic characteristics of the RR and non-RR Year 1 and Year 3 regression cohorts

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

	Year 1 Reading Texts Regression Sample		Year 3 NAPLAN Reading Regression Sample	
	RR (n = 3,187)	Non-RR (n = 3,408)	RR (n = 4,408)	Non-RR (n = 6,499)
Gender (%)				
Female	39.8	49.1	40.6	49.3
Male	60.2	50.9	59.4	50.7
Aboriginal Status (%)				
Non-Aboriginal	87.8	93.8	87.5	92.9
Aboriginal	10.1	3.8	10.1	4.3
Unknown	2.1	2.4	2.5	2.8
NESB (%)				
No	65.1	75.7	69.1	76.8
Yes	34.9	24.3	30.9	23.2
ESL Phase^a (%)				
Phase 1	15.6	6.3	14.0	5.5
Phase 2	2.4	3.9	2.3	3.1
Phase 3	0.5	1.0	0.5	0.8
Other ^a	81.5	88.8	83.3	90.7
Parental Education Level (%)				
Not stated	6.8	4.1	6.7	4.4
Year 9 or equivalent	13.5	3.6	12.7	3.8
Year 10 or equivalent	29.5	18.1	30.0	18.0
Year 11 or equivalent	9.0	5.6	9.2	5.8
Year 12 or equivalent	41.2	68.6	41.4	68.0
Parental Occupation Group (%)				
Senior Management	4.8	18.8	5.3	19.1
Other Business Manager	8.6	19.2	9.3	19.5
Trades/Clerk/Sales	14.5	18.1	15.9	17.6
Machine Operator/Hospitality Staff	14.1	10.3	14.4	10.5
Not in paid work for 12 months	36.9	23.5	34.8	22.6
Not stated	21.1	10.0	20.3	10.7

Note. All differences between the RR and non-RR cohorts included in each regression are statistically significant at $p < .001$. Demographic characteristics for gender, Aboriginal Status, NESB and ESL Phase were measured at Term 1, Kindergarten as part of the Best Start assessment. Parental education and occupation characteristics were measured at Year 3 as part of NAPLAN.

^aOther includes: not required, to be assessed, unknown and not provided.

11 The distribution shown for the end of Year 1 sample represents the students included in the regression for the Literacy Continuum aspect for Reading Texts. While there was a slight variation in the number of students included in each of the remaining Literacy Continuum aspect regressions, the same pattern of results presented in Table 1 was observed across all other Literacy Continuum aspects.

Table 2:**Distribution of Best Start scores for the RR and non-RR Year 1 and Year 3 regression cohorts**

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

Table 2 shows the distribution of Best Start scores for the final groups of RR and non-RR students included in the regression analyses of student outcomes at the end of Year 1 and at NAPLAN in Year 3¹². As shown in Table 2, compared to RR students, non-RR students in both regression samples were significantly more likely to be assessed at higher levels on all Best Start aspects at the beginning of Kindergarten (all p-values < .001). The largest discrepancy between RR and non-RR students across both regression samples was observed for Phonics, where over twice as many RR students were assessed at the lowest level (Level 0) compared to non-RR students.

	Year 1 Reading Texts Regression Sample		Year 3 NAPLAN Reading Regression Sample	
	RR (n = 3,187)	Non-RR (n = 3,408)	RR (n = 4,408)	Non-RR (n = 6,499)
Reading Texts (%)				
Level 0	81.3	49.1	78.8	48.6
Level 1+	18.7	50.9	21.2	51.4
Comprehension (%)				
Level 0	73.8	43.5	72.1	42.7
Level 1+	26.2	56.5	27.9	57.3
Aspects of Writing (%)				
Level 0	91.5	77.6	90.8	77.6
Level 1+	8.5	22.4	9.2	22.4
Aspects of Speaking (%)				
Level 0	57.9	24.4	54.3	24.2
Level 1+	42.1	75.6	45.7	75.8
Phonics (%)				
Level 0	76.5	33.8	74.6	32.1
Level 1+	23.5	66.2	25.4	67.9
Phonemic Awareness (%)				
Level 0	88.0	55.3	86.6	54.9
Level 1+	12.0	44.7	13.4	45.1
Concepts about Print (%)				
Level 0	82.4	47.3	80.6	46.3
Level 1+	17.6	52.7	19.4	53.7

Note. All differences between the RR and non-RR cohorts included in each regression are statistically significant at $p < .001$. Reported percentages exclude students with missing scores (less than 2% of each regression sample).

12 As above, the distribution shown for the end of Year 1 sample represents the students included in the regression for the Literacy Continuum aspect for Reading Texts. While there was a slight variation in the number of students included in each of the remaining Literacy Continuum aspect regressions, the same pattern of results presented in Table 2 was observed across all other aspects.

Modelling the Impact of RR on Year 1 Outcomes

To examine the impact of RR at the end of Year 1, a series of random effects regression models were used to compare Literacy Continuum scores at the end of Year 1 (Term 4) for the RR and non-RR groups¹³. Student outcomes were examined on seven aspects of the Continuum, including: Reading Texts; Comprehension; Aspects of Writing; Aspects of Speaking; Phonics; Phonemic Awareness; and Concepts about Print. Each aspect was treated as an individual outcome and separate models were fitted for each of the seven aspects¹⁴.

The Literacy Continuum scoring structure was not designed as an interval scale. Therefore, a series of random effects ordered logistic regression models (also known as proportional odds models) were fitted for each aspect. This type of approach assumes an ordinal structure to the Continuum, where the clusters are assumed to be sequentially ordered, but not necessarily equally spaced. In each model, the dependent variable was the level (cluster) achieved by the student on the relevant Literacy Continuum aspect at the end of Year 1¹⁵.

The same independent variables were used in each model to control for relevant student- and school-level factors known to be associated with academic performance. These included: gender, Aboriginal status, country of birth, NESB, ESL Phase at Best Start (Term 1 Kindergarten), parent qualification, parent occupation group, parent's spoken language at home, school participation in the Priority Schools Funding Program (PFSP) at Term 4 Kindergarten, school location at Term 4 Year 1, Family Occupation and Education Index (FOEI) at Term 4 Year 1¹⁶, and school region at Term 4 Year 1. In addition, the corresponding Literacy Continuum aspect level achieved by each student at Best Start and Term 4 Kindergarten were included in each model to account for important baseline differences in RR and non-RR students' literacy skills prior to the availability of RR in Year 1.

Two interaction terms were included in each model: (1) Literacy Continuum aspect level at Term 4 Kindergarten and treatment group (RR vs. non-RR); and (2) Literacy Continuum level at Best Start and treatment group (RR vs. non-RR). The interaction terms were included to assess whether there were any differences between RR and non-RR students as a function of their baseline performance prior to Year 1.

Students were included in the regression model for each aspect if they had non-missing demographic information for the independent variables and valid scores (i.e. not missing, blank or unknown) for the relevant aspect in the Best Start assessment in 2011 and in the Literacy Continuum assessments at the end of Term 4 in Kindergarten (2011) and at the end of Year 1 (2012). If these criteria were not met, the student was excluded from the analysis. Since assessment on the Continuum is not mandatory after Best Start, this resulted in a large number of missing records at the end of Kindergarten and at the end of Year 1 for both RR and non-RR students. Consequently, among the students who had matched data records and were eligible for inclusion in the regressions (7,208 RR students and 13,321 non-RR students), approximately 56 per cent of the RR sample and nearly 75 per cent of the non-RR sample were excluded in each regression. In all analyses, p-values of less than 0.05 indicate that the odds ratio is significantly different from one. Appendix B provides a further description of the modelling approach, including relevant model assumptions and diagnostics.

Since students were not randomly assigned to participate in RR, it is possible that the above regression analyses were influenced by unmeasured confounding factors that were associated both with the likelihood of participating in RR and Literacy Continuum outcomes at the end of Year 1 (i.e. omitted variable bias). To address this concern, a propensity score matching (PSM) approach was used to assess the robustness of the findings from the random effects regressions. In brief, the PSM approach attempts to minimise bias by approximating the conditions of randomisation for matched pairs of students that are comparable on all observed characteristics. Following matching, student outcomes at the end of Year 1 are compared for all matched pairs. Appendix C gives a detailed description of the PSM approach.

13 Random effects models were selected as they are suitable for fitting nested data, as is the case with assessing student performance within schools. In the model, the random school-specific effects can control for any unobserved heterogeneity across schools (that is constant across students within a school and is uncorrelated with the independent variables used in the model). In the random effects model, schools are considered as a sample drawn from a wider population of schools, and the school-specific effects (u_{0j}) are assumed to be random. The population is considered to have a hierarchical structure with level 1 units (students) clustered within level 2 units (schools).

14 Each aspect on the Literacy Continuum represents a specific skill-set and it may not be appropriate to aggregate (i.e. sum) scores across all outcomes to yield a total Continuum score for the regression analyses.

15 In all subsequent analyses, clusters are treated as levels, with Level 0 equal to Cluster 1; Level 1 equal to Cluster 2 etc.

16 FOEI is measure of socio-economic status developed by the Centre for Education Statistics and Evaluation (CESE 2014).

Impact of RR on Year 3 Reading Outcomes (NAPLAN Analysis)

A random effects linear regression model was used to examine whether there were any longer-term effects of RR on the scaled NAPLAN Reading scores achieved in Year 3 in 2014. The NAPLAN Reading score is an interval variable and is suitable for analysis using a linear regression approach. In this model, the dependent variable was the scaled NAPLAN Reading score in Year 3. The independent variables in the model included scores (i.e. levels) on the Literacy Continuum aspects for Reading Texts and Comprehension (as assessed at Best Start and Term 4 Kindergarten)¹⁷, which were included to control for potential baseline differences between RR and non-RR students' literacy skills prior to the availability of RR in Year 1. The model also included the same set of student- and school-level independent variables that were included in the ordered logistic regression models described above. Finally, an interaction term corresponding to the Reading Texts level achieved at Term 4 Kindergarten and treatment group (RR vs. non-RR) was included in the model to assess the effect of RR for students at different baseline starting levels prior to the availability of RR.

Students were included in the regression if they had non-missing demographic information for the independent variables, a valid (i.e. non-missing) NAPLAN score on the Reading domain in Year 3, as well as valid scores on the Literacy Continuum aspects for Reading Texts and Comprehension at Best Start and at Term 4 Kindergarten. If these criteria were not met, the record was excluded from the analysis. Among the students who had matched data records and were eligible for inclusion in the regression (7,208 RR students and 13,321 non-RR students), approximately 39 per cent of the RR sample and 51 per cent of the non-RR sample were excluded, leaving a final regression sample of 10,907. For each estimated effect, a p-value of less than 0.05 indicates that the effect is significantly different from zero. Appendix D provides a further description of the modelling approach, including relevant model assumptions and diagnostics.

As above, the robustness of the random effects linear regression to the impact of any potential omitted variable bias was assessed using PSM (see Appendix C).

17 Data corresponding to student progress on the Continuum at the end of Year 1 was not included in the NAPLAN regression.

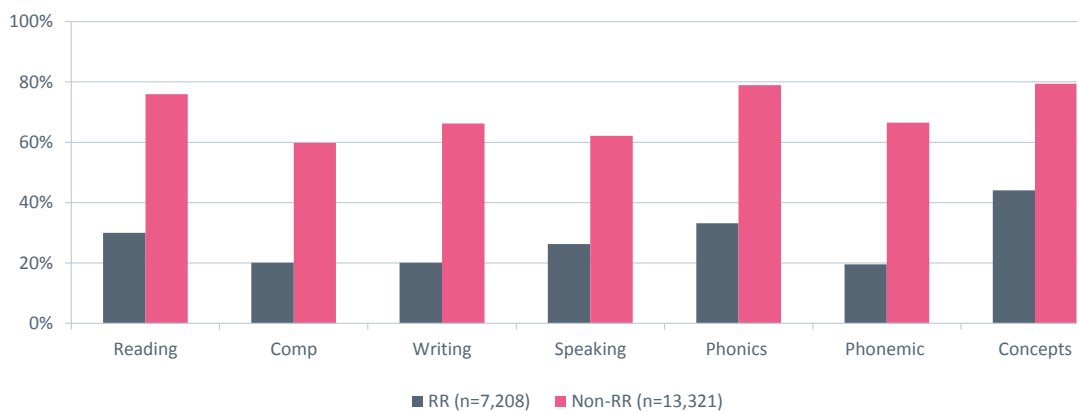
3. Results

Examination of student outcomes at the end of Kindergarten (prior to the availability of RR) revealed large differences in the proportion of RR and non-RR students meeting expected standards on the Literacy Continuum¹⁸. As shown in Figure 2, a significantly lower proportion of RR students were assessed as having met expectations across all Continuum aspects at the end of Kindergarten relative to their non-RR counterparts (all p-values < .05). The largest discrepancies between RR and non-RR students meeting expected standards were observed for Phonemic Awareness (19.6% vs. 66.5%), Aspects of Writing (20.1% vs. 66.2%), Reading Texts (30.0% vs. 76.0%) and Phonics (33.1% vs. 78.9%).

Figure 2:

Proportion of RR and non-RR students 'meeting expectations' on the Literacy Continuum at Term 4, Kindergarten

Sources: Reading Recovery and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.



Note. For each aspect, students were excluded from reported percentages if their scores were missing or listed as non-assessment. The proportion of cases missing for each aspect combined across RR and non-RR cohorts was approximately 42.5%.

These findings are not surprising, as students who went on to participate in RR in Year 1 were selected on the basis of showing poor performance in reading. Nevertheless, there were a number of students who were assessed as meeting expectations on the Literacy Continuum at the end of Kindergarten who went on to participate in RR. For example, 30.0% of students who participated in RR in Year 1 met expected benchmarks on Reading Texts at the end of Kindergarten. This suggests some lack of alignment between the screening measures used for RR and the current Literacy Continuum benchmarks.

¹⁸ At the end of Kindergarten, students are expected to reach Cluster 4 (Level 3) on all aspects. Scores on the vocabulary aspect are excluded from analysis as this aspect is not assessed at Best Start.

What proportion of students participating in RR reach the minimum reading levels expected of Year 1 students, and achieve literacy outcomes equal to or greater than those of their peers?

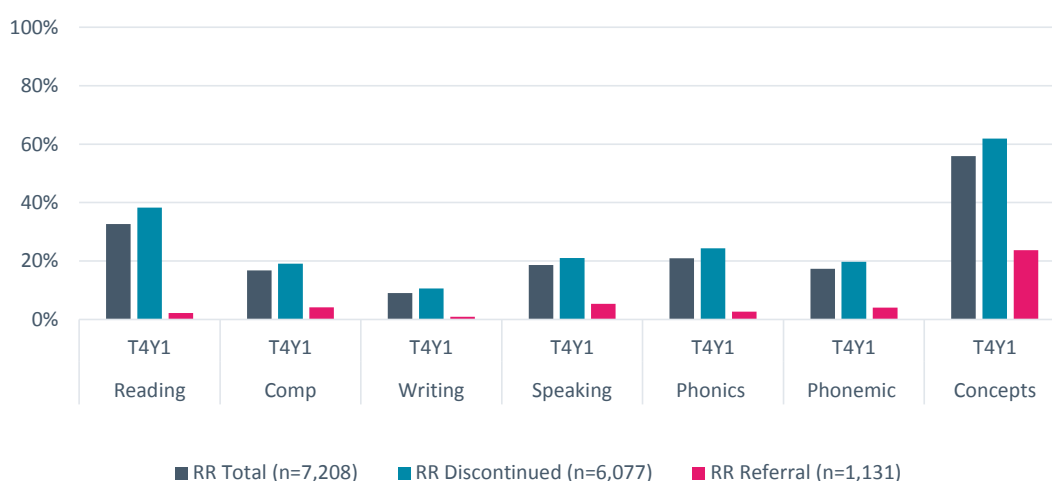
For students participating in RR in 2012, program outcomes show that 80.2% of students successfully discontinue (i.e. complete RR), 14.9% are referred for further specialist or long-term literacy support, while the remaining 4.9% either transfer out of the program or have their lessons carried over to the following year¹⁹.

However, the high program success rate contrasts starkly with the proportion of RR students who were assessed by teachers as having met the expected standards on Literacy Continuum measures following participation in RR at the end of Year 1. As shown in Figure 3, among students who participated in RR (discontinued and referred students only) and had valid Continuum scores at the end of Year 1, the majority did not meet Continuum expectations²⁰. This was true for both students who were successfully discontinued and those who were referred for long-term literacy support. However, a higher proportion of those that were discontinued from RR were assessed as meeting expectations compared to those referred for further support. Among the total cohort of discontinued and referred RR students, the proportion of students who met expectations ranged from 9.1% for aspects of writing to 55.9% for concepts about print. For Reading Texts, only 32.7% of RR students were deemed by their classroom teachers to be meeting expected standards at the end of Year 1.

Figure 3:

Proportion of RR students 'meeting expectations' on the Literacy Continuum at Term 4, Year 1

Sources: Reading Recovery and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.



Note. For each aspect, approximately 40.1% of students had missing scores or were listed as non-assessment. These students were excluded from reported percentages.

Figure 4 compares the proportion of students in RR and non-RR groups meeting expectations on the Literacy Continuum at the end of Year 1. A significantly lower proportion of RR students (discontinued and referred) met expectations compared to their non-RR counterparts (all p-values < .001). In 2012, more than twice as many non-RR students met expected standards for Reading Texts compared to RR students (67.4% vs. 32.7%), while for Phonics and Phonemic Awareness, roughly three times as many non-RR students met expectations compared to RR students (63.9% vs. 21.0% and 55.5% vs. 17.3%, respectively).

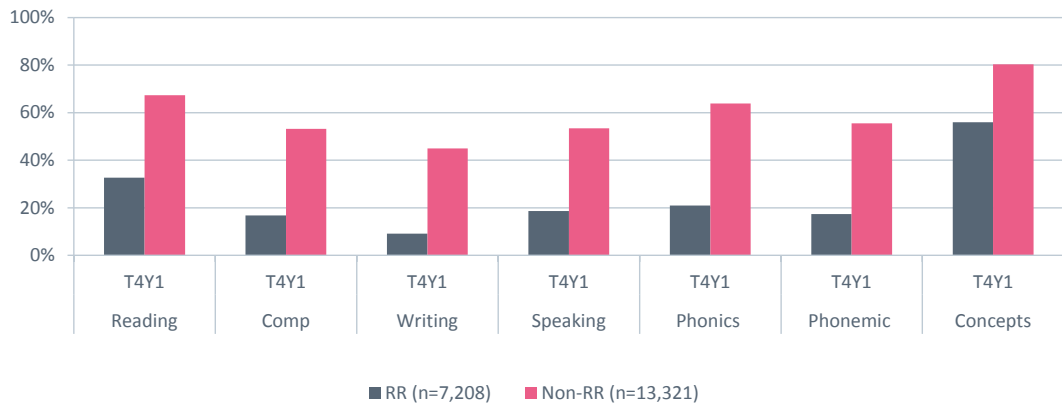
¹⁹ This includes all students who participated in RR in 2012 and had matched data records from the Best Start assessment in 2011, the Literacy Continuum assessments at the end of Kindergarten (2011) and at the end of Year 1 (2012), and at Year 3 NAPLAN (n = 7,573).

²⁰ At the end of Year 1, students are expected to reach Cluster 6 on all aspects except concepts about print, where expectations are at a Cluster 5. Scores on the vocabulary aspect are excluded from analysis as this aspect is not assessed at Best Start.

Figure 4:

Proportion of RR and non-RR students 'meeting expectations' on the Literacy Continuum at Term 4, Year 1

Sources: Reading Recovery and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.



Note. For each aspect, students were excluded from reported percentages if their scores were missing or listed as non-assessment. The proportion of cases missing for each aspect combined across RR and non-RR cohorts was approximately 55%.

This pattern of results is similar when looking specifically at the Literacy Continuum aspect for Reading Texts and examining differences in outcomes while accounting for baseline levels of achievement at the end of Kindergarten. As presented in Figure 5, the gap between RR and non-RR students meeting Year 1 expectations in Reading Texts in 2012 was smaller for students who had lower starting points at the end of Kindergarten compared to those with higher starting points. However, the overall pattern of results shows that non-RR students were still more likely to have met Year 1 reading expectations, irrespective of their prior year's progress on the Continuum.

Figure 5:

Proportion of students 'meeting expectations' on Reading Texts at Term 4 Year 1, by Term 4 Kindergarten scores

Sources: Reading Recovery and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.



Note. Student data was excluded from reported percentages if scores were missing or listed as non-assessment at either Term 4 Kindergarten or Term 4 Year 1. The proportion of cases missing across both cohorts was 66.8% (54.7% missing for the RR group and 73.4% missing for the non-RR group).

The above results comparing student outcomes against expectations on the Literacy Continuum should be treated with some degree of caution. The Continuum was developed as a tool to assist teachers to understand and describe the development of literacy skills among students. It was not designed as a reliable and valid measure of student progress against robust benchmarks. Indeed, following Best Start, teachers make subjective judgements about student progress on the Continuum by extrapolating the outcomes from school-based assessments and classroom observation to align with the Continuum clusters. The magnitude of the discrepancy between RR benchmarks and Continuum outcomes suggests that the expected standards may require some further refinement. Nevertheless, it is clear from the data presented above that RR students are less likely to meet Continuum expectations and show lower growth than their non-RR counterparts (in non-RR schools) at the end of Year 1.

It is important to note that these findings do not indicate that RR is an ineffective intervention. The effectiveness of RR can only be ascertained through more rigorous analysis that controls for the impact of important contextual factors and key differences in the baseline literacy skills of RR and non-RR students prior to Year 1. This analysis is summarised in the next section.

In the short-term, are literacy outcomes for students who participate in RR greater than those for comparable students who do not participate in RR?

To test the effectiveness of RR on Literacy Continuum outcomes at the end of Year 1, random effects regression models were constructed controlling for the impact of important student- and school-level factors. The results for each of the seven Literacy Continuum aspects are presented as odds ratios (see Tables 3 and E1), which are defined as the odds of reaching a higher level in the Literacy Continuum for a given category (e.g. gender male) relative to the baseline category (e.g. gender female). An odds ratio that is significantly greater than 1 indicates that students in the corresponding category have higher odds of progressing to a higher level in the Literacy Continuum than students in the baseline category. An odds ratio significantly less than 1 indicates that students in the corresponding category have lower odds of progressing than students in the baseline category.

There was a significant interaction between participation in RR and students' baseline outcomes at the end of Kindergarten. As shown in Table 3, students who participated in RR and who were assessed at Level 1 or below in Reading Texts at Term 4 Kindergarten had odds of progressing to a higher Reading Texts level at the end of Year 1 that were 2.7 times higher than their non-RR counterparts. The magnitude of this effect decreased to 1.5, and was not statistically significant for students who started at Level 2 in Reading Texts at the end of Kindergarten. Above Level 2, the odds ratios for Reading Texts (0.4 Level 3 and 0.09 Level 4 or above) reveal that non-RR students had much higher odds of progressing than RR students. For the remaining Literacy Continuum aspects, no other significant positive effects of RR were observed. Indeed, all other significant effects in Table 3 represent odds ratios less than 1, showing that non-RR students at higher starting points at the end of Kindergarten had better short-term outcomes than RR students, even after accounting for important contextual variables.

Table 3:

Estimated effect of RR on the Literacy Continuum level achieved at the end of Year 1

Sources: Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

Lit. Cont. Aspect	Lit. Cont. Aspect Level (T4K)	Odds Ratio	Lit. Cont. Aspect	Lit. Cont. Aspect Level (T4K)	Odds Ratio
Reading Texts	Level 1 or below	2.67*	Phonics	Level 1 or below	0.96
	Level 2	1.50		Level 2	1.01
	Level 3	0.42*		Level 3	0.28*
	Level 4 or above	0.09*		Level 4 or above	0.14*
Comprehension	Level 1 or below	1.3	Phonemic Awareness	Level 1 or below	0.85
	Level 2	0.53*		Level 2	0.52*
	Level 3	0.30*		Level 3	0.29*
	Level 4 or above	0.13*		Level 4 or above	0.22*
Aspects of Writing	Level 1 or below	0.98	Concepts about Print	Level 1 or below	1.11
	Level 2	0.40*		Level 2	0.53*
	Level 3	0.22*		Level 3	0.33*
	Level 4 or above	0.11*		Level 4 or above	0.05*
Aspects of Speaking	Level 1 or below	0.78			
	Level 2	0.51*			
	Level 3	0.44*			
	Level 4 or above	0.34*			

Note. For each Literacy Continuum aspect, the odds ratio is estimated for each aspect level in Term 4 Kindergarten. It represents the odds ratio of achieving a higher level in Literacy Continuum at the end of Year 1 for RR students relative to non-RR students. *Indicates that the odds ratio is significantly different from one at the .05 significance level. Results coloured in **cyan** show that RR students had higher odds of progressing on the Continuum relative to non-RR students; results coloured in **pink** show that non-RR students had higher odds of progressing on the Continuum relative to RR students.

As shown in Appendix E (see Table E1), other significant effects that were consistent across most aspects at the end of Year 1 revealed that, after controlling for all other factors in the model:

- Male students had lower odds of progressing on the Continuum compared to female students.
- Aboriginal students had lower odds of progressing on the Continuum than non-Aboriginal students.
- Students born in Australia and external territories had lower odds of progressing on the Continuum than students born outside Australia.
- Students whose parents had achieved a Bachelor degree or above had higher odds of progressing on the Continuum relative to students whose parent qualification was not stated.
- Students who attended primary schools in the former Illawarra and South East region in Year 1 had higher odds of progressing on the Continuum relative to students who attended schools in Sydney.

In addition, PSM was used to test whether the above findings were impacted by any bias associated with unmeasured variables that could influence both the likelihood of receiving RR and Continuum outcomes at the end of Year 1. Results (reported in Table E2) showed a very similar pattern to that presented in Table 3. Students who participated in RR and were assessed at the lowest levels on Reading Texts at the end of Kindergarten had odds of progressing to a higher Reading Texts level at the end of Year 1 that were 1.9 times higher than their non-RR counterparts. However, for all other aspects the results were equivalent for RR and non-RR students at the lowest baseline starting points. For students assessed at higher levels on the Continuum at the end of Kindergarten, results showed that non-RR students had higher odds of progressing relative to RR students.

Sensitivity Analysis

To test the robustness of the modelling results presented in Table 3, sensitivity analyses were undertaken examining students' Literacy Continuum outcomes at the end of Year 1. In the first analysis, students who were referred for long-term literacy support were excluded from the RR cohort to account for the possibility that some of these students may have suffered from a learning disability or other cognitive impairment that could have confounded their achievement outcomes following RR. The pattern of results from this analysis (summarised in Table E3) are similar to those presented in Table 3. The major difference was in the magnitude of the RR effects, particularly for students with lower baseline levels of achievement at the end of Kindergarten, which increased for all Continuum aspects. With the exception of Aspects of Speaking, students who successfully discontinued RR and were assessed at Level 1 or below on each relevant Continuum aspect at the end of Kindergarten had significantly higher odds of progressing than their non-RR counterparts (all p-values < .05). For example, for Reading Texts, RR students who were assessed at Level 1 or below at the end of Kindergarten had odds of progressing on this aspect that were 7.8 times greater than non-RR students. This represents a substantial increase in the effect of RR that was observed when both discontinued and referred students were included in the RR cohort (odds ratio of 2.7, as shown in Table 3).

The magnitude of the RR effect in the sensitivity analysis remained statistically significant for students who started at Level 2 in Reading Texts at the end of Kindergarten (odds ratio of 2.1). For students at higher starting points at the end of Kindergarten (i.e. Levels 3 and 4), results were consistent with those presented in Table 3 showing that non-RR students had higher odds of progressing than RR students. Given that it is unlikely that all students referred for long-term literacy support suffered from learning disabilities, and the fact that some students in the non-RR group would also have had learning difficulties, the true effect of RR on student outcomes at the end of Year 1 is likely to lie at some point between the estimates summarised in Table 3 (including all discontinued and referred students) and those observed in Table E3 (including discontinued students only).

A second sensitivity analysis was conducted to account for the potential impact of student mobility across schools on teachers' Literacy Continuum assessments at the end of Year 1. Examination of the distribution of students in each cohort who changed schools between the end of Kindergarten and the end of Year 1 showed that approximately 7 per cent of the RR cohort and 4 per cent of the non-RR cohort moved schools during this period. Within each cohort, students who moved schools during this period were more likely to show negative or zero growth on the Continuum aspects compared to students who stayed in the same school. These findings suggest that teachers from the same school are more likely to make judgements on the Continuum that show progress between the end of Kindergarten and the end of Year 1 than teachers at different schools. The results of this sensitivity analysis (summarised in Table E4) revealed a very similar pattern to that displayed in Table 3 (i.e., an odds ratio of 2.8 for the effect of RR on Reading

Texts for students assessed at Level 1 or below at the end of Kindergarten). These results suggest that the modelling outcomes presented in Table 3 are robust to the impact of any potential variation in teachers' assessments of growth on the Continuum within and across schools.

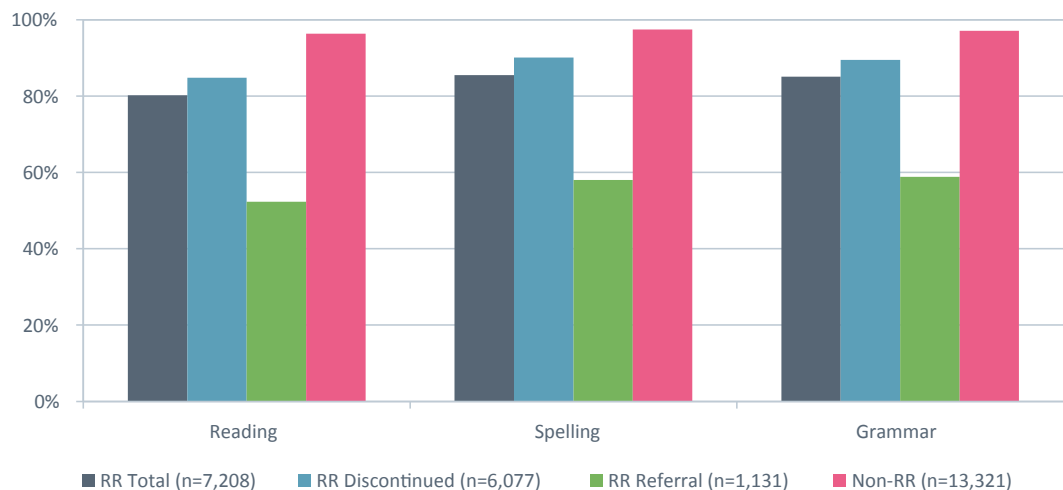
Are any benefits of RR sustained over longer periods of time?

Year 3 NAPLAN scores from 2014 were analysed to examine whether there were any longer-term differences between RR and non-RR cohorts. This analysis begins with unadjusted comparisons of the attainment of National Minimum Standards in Year 3 NAPLAN for RR students and non-RR students, followed by a breakdown of Year 3 Reading scores by baseline reading level (as assessed on Reading Texts at the end of Kindergarten). The final analysis examines differences in NAPLAN reading scores for RR and non-RR students, accounting for differences in students' baseline literacy levels as well as other important student- and school-level characteristics. This final analysis provides the best indication of the sustained effects of RR.

Figure 6 shows that a significantly higher proportion of non-RR students achieved the National Minimum Standards (Band 2 or above) in Year 3 (2014) Reading, Spelling and Grammar, compared to students who participated in RR (all p-values < .001). Across domains, the gap between non-RR students and all RR students ranged between 12% (Spelling and Grammar) to 16% (Reading). When outcomes for RR students were broken down by program exit status and compared to non-RR students, the results showed a smaller gap for RR students who successfully discontinued the program and a much larger gap for those who were referred for longer-term literacy support.

Figure 6:
Proportion of RR and non-RR students meeting National Minimum Standards on NAPLAN Year 3 assessments²¹

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery data sourced from the Early Learning and Primary Education Directorate, DOE.



Note. Student data was excluded from reported percentages if NAPLAN scores were missing. The proportion of cases missing was roughly similar across domains (approximately 5% of the RR cohort and 3% of the non-RR cohort).
*Differences between RR students (all, discontinued, referred) versus non-RR students were significant at $p < .001$ across all domains.

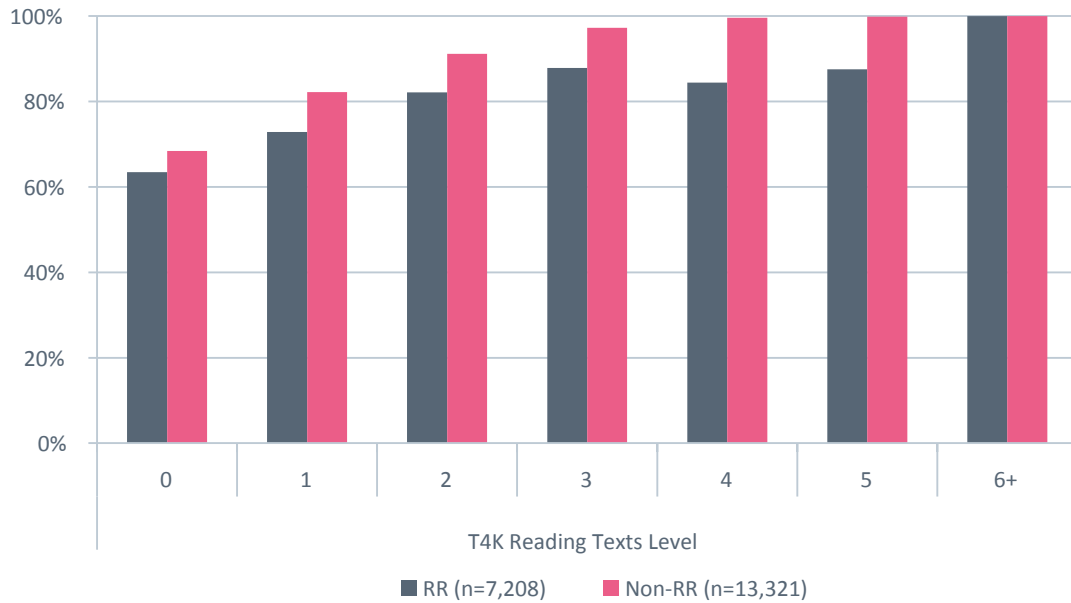
A similar pattern of results emerged when student outcomes on the NAPLAN Year 3 Reading domain were broken down by baseline reading level at the end of Kindergarten (as assessed using the Reading Texts aspect of the Literacy Continuum at Term 4 Kindergarten). As shown in Figure 7, except for the highest level, a higher proportion of non-RR students met National Minimum Standards in Year 3 NAPLAN (2014) compared to RR students.

²¹ Outcomes for the Writing Domain of the NAPLAN Year 3 (2014) assessment are not included due to concerns that the nature of the task may have been too difficult, especially for primary-aged students. In NSW, the number of students who did not attempt the task and who subsequently received a zero score substantially increased in 2014 compared to previous years.

Figure 7:

Proportion of RR and non-RR students meeting National Minimum Standards on NAPLAN Year 3 Reading, by Term 4 Kindergarten Reading Text scores

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.



Note. Student data was excluded from reported percentages if scores were missing at either Term 4 Kindergarten or NAPLAN. The proportion of cases missing across both cohorts was 44.6% (36.7% missing for the RR cohort and 48.8% missing for the non-RR cohort).

As with the analysis of Continuum data, regression analyses were also undertaken to investigate the effect of RR on NAPLAN Reading scores whilst controlling for students’ baseline literacy skills and other relevant student- and school-level factors. The results of the regression are presented as regression coefficients (see Table 4 and Appendix F (Table F1)), which represent the difference in the scaled Year 3 NAPLAN Reading score for those in the corresponding category (e.g. gender male) relative to the baseline category (e.g. gender female) while holding all the other independent variables constant. Positive coefficients are associated with higher NAPLAN scores in Year 3 Reading, while negative coefficients reflect lower scores. For example, a regression coefficient of -3.1 for male students reveals that males had NAPLAN Reading scores that were, on average, 3.1 points lower compared to female students.

There was a significant interaction between participation in RR and students’ baseline reading levels on NAPLAN outcomes. As shown in Table 4, students who participated in RR achieved lower scores on the Year 3 NAPLAN Reading domain irrespective of their starting points but the effect was much larger for students at higher starting points. For students at the lowest levels at the end of Kindergarten (Level 1 or below), the average difference in scaled NAPLAN Reading scores was 25.2 points in favour of non-RR students. For students at the highest levels at the end of Kindergarten (Level 4 or above), the difference in Year 3 NAPLAN Reading scores increased to 86.7 points in favour of non-RR students.

Table 4:

Estimated effect of RR on NAPLAN Reading outcomes in Year 3

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

Reading Texts level at T4K	RR effect on NAPLAN Reading score	p-value
Level 1 or below	-25.2	<.001
Level 2	-24.9	<.001
Level 3	-53.1	<.001
Level 4 or above	-86.7	<.001

Note. Results coloured in pink show that RR students achieved lower NAPLAN Reading scores compared to non-RR students.

Other significant effects (summarised in Table F1) revealed that, holding all other variables in the model constant:

- Aboriginal students scored 16.0 points lower than Non-Aboriginal students.
- Students born in Australia and other external territories scored 6.8 points lower than students born outside Australia.
- Students who were at ESL Phase 2 or Phase 3 at the Best Start assessment achieved lower scores compared to students in ESL Phase 1 (9.2 and 18.1 points, respectively).
- Students whose parents achieved an Advanced Diploma/Diploma or a Bachelor degree achieved higher scores compared to those whose parents qualification was not stated (9.0 and 20.9 points, respectively).
- Students who attended schools that were more disadvantaged (had higher FOEI scores) achieved lower scores compared to students from more advantaged schools.
- Students who attended primary schools in the former Illawarra and South East region in Year 1 scored 10.5 points lower than students who attended schools in Sydney.

In addition, PSM was used to test whether the above findings were robust to other modelling specifications. Results (reported in Table F2) showed a very similar pattern to that presented in Table 4, with all effects pointing towards higher NAPLAN scores for non-RR students, irrespective of their baseline starting points.

Sensitivity Analysis

To test the robustness of the modelling results examining Year 3 NAPLAN outcomes, a sensitivity analysis was undertaken excluding students referred for long-term literacy support from the RR cohort, as described above. Results from this analysis showed that students who successfully discontinued from RR still achieved significantly lower scores on the Year 3 NAPLAN Reading domain irrespective of their baseline Reading Texts outcomes at the end of Kindergarten. However, the observed effects for students at the lowest levels at the end of Kindergarten were smaller than those described in Table 4 (see Table F3). For example, for students at Level 1 or below on Reading Texts at the end of Kindergarten, the average difference in scaled NAPLAN Reading scores was halved to 12 points (from 25.2) in favour of non-RR students. For students at higher starting points at the end of Kindergarten, the difference between the results of the sensitivity analysis and those observed in Table 4 was only 4 points for students at Level 2 and 1 point for students at Level 3 and above. As noted in the analysis of student outcomes at the end of Year 1, the results of this sensitivity analysis suggest that the true effect of RR on student outcomes in Year 3 is likely to lie between the outcomes summarised in Table 4 including all discontinued and referred students and those including discontinued students only (Table F3).

4. Discussion

The results from this retrospective evaluation provide some evidence that RR is effective at improving short-term reading outcomes at the end of Year 1. However, this was only evident on the Continuum aspect corresponding directly to Reading Texts and was only observed for students identified as the lowest performing readers at the end Kindergarten. For students achieving higher reading levels at the end of Kindergarten, RR was significantly less effective at improving reading performance at the end of Year 1 compared to the instruction delivered to non-RR students in non-RR schools. These findings suggest that RR is an effective intervention for improving short-term reading outcomes among the poorest performing readers, which is the primary intention of the intervention.

These findings sit in contrast to the criticisms of the program raised in New Zealand, where it has been argued that RR is least effective for those at the lower end of the achievement spectrum (Tunmer et al. 2013). The current results are consistent with earlier findings from robust studies summarised in the What Works Clearinghouse (2013) review and in the analysis conducted in NSW by Center et al. (1995). In the What Works Clearinghouse (2013) review, the most robust effects of RR were observed on general reading achievement measures assessed shortly following the completion of RR. In relation to the remaining literacy aspects on the Continuum, the results for the lowest performing readers were equivalent for RR and non-RR students. However, for those at higher starting points at the end of Kindergarten, results showed that non-RR students made significantly greater progress on the Continuum measures than RR students. These findings suggest that, for students at higher baseline starting literacy levels, RR is not as effective as the instruction and interventions offered at non-RR schools.

In relation to the relative longer-term intervention effects, results showed no evidence that RR yields any positive effects on students' NAPLAN Reading performance in Year 3. Irrespective of students' starting ability, students who participated in RR achieved significantly lower scores on the Year 3 NAPLAN Reading assessment compared to their non-RR counterparts. The lack of sustained benefit associated with RR should not necessarily be interpreted as a failure of the RR program. The duration of the program is only 12-20 weeks so it is equally possible that RR students do not receive the level of support they need to sustain any short-term RR effects beyond Year 1. There is a relatively large time lag between student participation in RR in Year 1 and NAPLAN testing in Term 2 of Year 3. Students, especially those identified as at-risk of falling behind, are often exposed to a range of initiatives and interventions throughout the early years of school.

While the current study provides the first rigorous sector-level analysis of the relative longer-term effect of RR in NSW, the impact of RR on students' Year 3 outcomes cannot be isolated from the potential effects of any exposure to other initiatives in the intervening period between Year 1 and Year 3. Furthermore, it cannot be assumed that schools offering RR would necessarily respond in the same way to remediating or building the literacy skills of students throughout the early years as schools that do not offer RR. Future research examining the longer-term impact of RR would benefit from capturing and controlling for students' exposure to other interventions or classroom initiatives that could impact the development of literacy skills to Year 3.

While the current findings reveal short-term positive effects of RR on reading outcomes for the lowest performing students, they do not support the effectiveness of the intervention on other aspects of literacy achievement or the longer-term sustainability through the early years of school. The reasons for these findings are not evident in the current study and will only be borne out by a closer inspection of the way RR is being delivered in NSW government schools. One possible explanation that is asserted strongly by RR critics is that RR does not provide sufficient tuition in phonics and phonemic awareness to effectively remediate literacy performance among struggling readers (Center et al. 1995; Chapman & Tunmer 2011; Greaney 2011; Moats 2007; Reynolds & Wheldall 2007; Tunmer & Chapman 2003; Tunmer et al. 2013). Center et al. (1995) argue that "while Reading Recovery stresses the importance of using all sources of information available to access meaningful text, it may not provide enough systematic instruction in the metalinguistic skills of phonemic awareness, phonological recoding, and syntactic awareness for students to acquire these processes" (p. 244).

It is well-accepted in the scientific literature that systematic early instruction in phonics is critical to the development of early reading skills, particularly for struggling students, and should form an integral component of a balanced approach to literacy development (Education Endowment Foundation 2015; Ehri et al. 2001; Hattie 2009; U.S. Department of Health and Human Services 2000; Rowe 2005). In Hattie's (2009) comprehensive analyses examining what works in lifting student achievement, phonics-based approaches achieved a moderate to high average effect size of 0.6, while whole language based approaches were shown to have a negligible effect on student performance (effect size = 0.06). Unfortunately, information is not available at the sector-level on what types of remedial literacy interventions are offered in non-RR schools. It may be the case that non-RR students received interventions that comprised more comprehensive instruction in phonics and were therefore more effective in remediating student literacy skills through to Year 3. This is an important empirical question that should be the focus of future research.

A potential factor that may have impacted the observed outcomes at the end of Year 1 could relate to teacher expectation and perceptual bias effects. Previous research suggests that teachers can develop expectations for student performance that influence their interactions with students and their assessments of student progress and achievement (Jussim 1989; Jussim, Eccles & Madon 1996; Rosenthal & Jacobson 1966). This raises the possibility that the process of identifying students as poor performing readers who require a withdrawal intervention such as RR may shape teachers' expectations of students' subsequent performance. That is, teachers may develop lower expectations of success for RR students, which could potentially bias the subsequent judgements that these teachers make for RR students on the Continuum at the end of Year 1. While this contention is speculative, it would at least in part explain why RR students at higher starting levels at the end of Kindergarten exhibit less growth on the Continuum at the end of Year 1 compared to non-RR students. For the RR students at lower starting points who show either greater or similar growth compared to non-RR students at the end of Year 1, it may be the case that the magnitude of the RR effect for these students is greater, or at least equivalent to any potential expectancy bias demonstrated by teachers. Whereas for students at higher starting points it could be the case that the impact of RR diminishes and the influence of potential teacher bias becomes more evident. However, even if this speculation holds true, it is important to bear in mind that RR students still perform more poorly on the Year 3 NAPLAN test, which is not dependent on teacher judgement and is therefore not subject to this bias.

Another possible factor that may limit the effectiveness of RR relates to teacher quality and intervention fidelity, neither of which could be assessed in the current analysis. A key assumption of RR is that high quality instruction is essential to accelerating the literacy skills of struggling students (May et al. 2013; 2015). That is, ensuring that RR is delivered by high quality teachers who can respond effectively and appropriately to each student's needs is critical to the success of the intervention. While the RR guidelines in NSW stipulate that all RR teachers are required to undergo training for one year in the implementation of RR and undertake ongoing professional development (Department of Education and Communities 2015), the possibility remains that the actual quality and consistency of implementation may vary across schools. Future research should include an in-depth analysis of intervention fidelity to better understand whether, and to what extent, these factors may influence student outcomes.

Finally, an important concern identified through the current study was the large discrepancy in student outcomes on the RR Observation Survey and the Literacy Continuum assessments both prior to, and following participation in RR. For example, analysis of Continuum data showed that approximately 30 per cent of students who go on to participate in RR in Year 1 are judged to be meeting expected learning outcomes on Reading Texts at the end of Kindergarten²². While the Continuum benchmarks may require some further refinement and could under-estimate students' needs for further literacy support, these findings could also point to some issues with the implementation of RR in NSW. Indeed, it is clear that at a sector-level, there are currently some relatively high-performing students participating in RR. While these students may be lower performing readers in a relatively high-achieving school or class context, they may not be appropriate candidates to participate in an intervention such as RR that was developed to target the lowest performing students. Consequently, it is possible that students at higher starting points may be disadvantaged by participating in a withdrawal program relative to receiving quality classroom instruction.

22 This includes only those students from the RR sample (n=7,208) who had a valid assessment score on Reading Texts at the end of Kindergarten (approximately 33% were excluded).

Our examination of student outcomes following participation in RR also revealed substantial differences in the proportion of students who were successfully discontinued from RR (80%) and the proportion who were assessed as meeting expectations on the Continuum at the end of Year 1 (e.g. 32.7% for Reading Texts)²³. While these measures were developed for different purposes and are administered by different teachers, this discrepancy may point to weaknesses in the criteria for discontinuing students from RR. The Observation Survey has been criticised by scholars for potentially over-estimating learning effects for RR students due to its close alignment to the content delivered in RR lessons (Grossen et al. 1997; Reynolds & Wheldall 2007). Further empirical research is required to better determine whether these assessments appropriately benchmark students' literacy needs and progress throughout the early years of school.

Limitations

While the current study represents a robust retrospective analysis assessing the sector-level impact of RR in NSW, there are some important limitations to be acknowledged. First, the Literacy Continuum was not developed as a robust assessment tool to measure student progress throughout the early years of school. The Continuum was developed to be used as a professional learning and teaching tool to help teachers understand and describe the development of literacy skills among students. Apart from the Best Start assessment, judgements about students' progress on the Literacy Continuum are subjectively made by teachers by extrapolating the outcomes from school-based assessments and classroom observation to align with the Continuum clusters. This raises a number of potential issues in relation to the reliability of these judgements. The extent to which school-based assessments align with the Continuum levels is currently unclear. The ability of teachers to consistently map school-based assessments to Continuum benchmarks is also currently unclear. In the absence of information about the reliability and validity of the Continuum judgements, there is a risk that the estimates reported in this study could be biased and inconsistent. In addition, as noted above, the subjective nature of these judgments could potentially make teachers vulnerable to certain biases in assessing student performance, particularly in relation to any potential stigma associated with a student classified as a low performing reader. However, while the Literacy Continuum may require some refinement to be used as a rigorous measure of student progress, the assumption made in the current analysis was that the variability in consistency of teacher judgement did not differ systematically across RR and non-RR schools.

A second concern in relation to the use of the Literacy Continuum is selection bias, largely related to missing data for Continuum assessments at the end of Kindergarten and at the end of Year 1. Following Best Start, use of the Continuum is not mandatory for schools and teachers. This resulted in the loss of a large number of cases in determining the RR and non-RR cohorts and in conducting regression analyses of both short and long-term student outcomes. Importantly, a larger proportionate loss was observed for non-RR students, as non-RR schools were less likely to complete Continuum assessments for students following Best Start through to the end of Year 1. While the current analysis was predicated on the assumption that RR and non-RR schools choose to use the Continuum for similar reasons that were unrelated to student outcomes, analysis of each of the sub-samples included in the regression analyses revealed some differences in the demographic and performance profiles compared to the full cohort of students who participated in the Best Start 2011 assessment (see Appendix A). For example, there was some evidence that the subset of non-RR students included in the regression samples were slightly more advantaged in their demographic and achievement profiles (as assessed at Best Start and on Year 3 NAPLAN Reading outcomes) compared to the total population of non-RR students who completed Best Start in 2011. In contrast, for the subset of RR students included in the regression samples, there was some evidence that the students included in the assessment of outcomes at the end of Year 1 had a slightly lower achievement profile compared to the total population of RR students participating in Best Start.

23 This only includes those students from the RR sample (n=7,208) who had a valid assessment score on Reading Texts at the end of Year 1 (approximately 40% were excluded).

While the current analysis accounted for all available and measurable demographic and prior achievement indicators, there were some variables that may be causally related to student outcomes that could not be measured and accounted for in determining the impact of RR. For example, it was not possible to measure any within-school factors (e.g. teacher quality, access to other classroom-based initiatives and interventions) that may have impacted student performance. Furthermore, it was not possible to account for all important student-level differences; the most notable exclusion being an indicator of students' learning disabilities, which is likely to have a significant impact on student achievement. This was addressed to some extent in sensitivity analyses excluding students referred for long term literacy support, which may have included at least some students with learning disabilities or other cognitive impairments. The outcomes of sensitivity analyses revealed larger positive effects of RR on student outcomes at the end of Year 1 and smaller differences in Year 3 outcomes between non-RR and RR students at the lowest starting points. This is likely to reflect the upper-bound of the effectiveness of RR. The exclusion of referred students from the RR cohort is only a proxy for learning disabilities and the over-riding assumption in the current analysis was that any important omitted variables were equally distributed across RR and non-RR cohorts and did not bias the observed outcomes. While this assumption was not testable in the current analysis, the consistency in the outcomes observed using propensity score matching lends further support to the robustness of the current findings to the impact of any bias associated with omitted variables.

Summary

The results from the current analysis provide some evidence that RR is an effective short-term intervention for remediating reading text skills among the lowest performing students. However, RR does not appear to be an effective intervention for students who begin Year 1 with more proficient literacy skills, at least compared to other interventions or initiatives that are available in non-RR schools. These findings do not necessarily call into question the validity of RR, as the intervention is designed specifically to remediate reading skills among the poorest performing students. However, the implication of these findings is that currently, the most cost-effective method of implementing RR in NSW may be to target only the students performing at the lowest levels at the end of Kindergarten (at a sector- not a school-level) or to restrict RR to schools that are identified as having a proportionately higher number of students who are not meeting performance benchmarks in Kindergarten or early Year 1.

The limitations of this analysis also highlight the strong need to collect better information on the teaching practices and interventions being offered in non-RR schools and to develop valid and reliable measures of students' literacy progress throughout the early years of school. Perhaps more importantly, these findings suggest that the true effect of the relative impact of RR can only be assessed through a rigorous prospective trial where RR is compared to other interventions, using random assignment to treatment condition and valid and reliable outcome measures. Indeed, only a rigorous prospective trial can effectively control for other potentially confounding factors (e.g. teacher quality, classroom instruction) that may influence student outcomes and can provide policy makers and school leaders with the critically needed information on which intervention is the most effective and cost-effective approach to literacy remediation.

References

- Anderson, JA 1984, 'Regression and ordered categorical variables', *Journal of the Royal Statistical Society Series B*, vol. 46, pp. 1-30.
- Australian Council for Educational Research 2013, *Literacy and numeracy interventions in the early years of schooling: A literature review*, Report to the Ministerial Advisory Group on Literacy and Numeracy.
- Center, Y, Wheldall, K, Freeman, L, Outhred, L, & Mcnaught, M 1995, 'An evaluation of Reading Recovery', *Reading Research Quarterly*, vol. 30, pp. 240–263.
- Centre for Education Statistics and Evaluation 2014, *Family Occupation and education index (FOEI) 2013*, technical report, report prepared by K Rickard & L Lu.
- Chapman, JW & Tunmer, WE 2011, 'Reading Recovery: Does it work?', *Perspectives on Language and Literacy, Fall*, pp. 21-24.
- Clay, M 2002, *An Observation Survey of Early Literacy Achievement*. 2nd edn, Heinemann, Portsmouth, N.H.
- D'Agostino, JF & Murphy, J 2004, 'A meta-analysis of Reading Recovery in United States Schools', *Educational Evaluation and Policy Analysis*, vol. 26, no. 1, pp 23-38.
- Department of Education and Communities 2015, *Reading Recovery Guidelines for 2015 and beyond*, Department of Education and Communities, viewed 27 March 2015, http://www.curriculumsupport.education.nsw.gov.au/earlyyears/reading_recovery/docs/rr_guidelin_15.pdf
- Education Endowment Foundation 2015, *Evidence and Data: Phonics*, viewed 27 March 2015, <http://educationendowmentfoundation.org.uk/toolkit/toolkit-a-z/phonics/>
- Ehri, LC, Nunes, SR, Stahl, SA & Willows, DM 2001, 'Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's Meta-Analysis', *Review of Educational Research*, vol. 71, no. 3, pp. 393-447.
- Greaney, K 2011, 'The multiple cues or "searchlights" word reading theory: Implications for Reading Recovery', *Perspectives on Language and Literacy, Fall*, pp. 15-19.
- Grossen, B, Coulter, G & Ruggles, B 1997, *Reading Recovery: An evaluation of benefits and costs*, viewed 22 March 2015, <http://www.dys-add.com/resources/ReadingMethods/ReadingRecoveryUnivofORStudy.pdf>
- Hattie, J 2009, *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge, New York.
- Jussim, L 1989, 'Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy', *Journal of Personality and Social Psychology*, vol. 57, no. 3, pp. 469-480.
- Jussim, L, Eccles, J & Madon, S 1996, 'Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy', *Advances in Experimental Social Psychology*, vol. 28, pp. 281-388.
- Long, JS & Freese, J 2006, *Regression models for categorical dependent variables using Stata*, 2nd edn. Stata Press, Texas.
- May, H, Gray, A, Gillespie, JN, Sirinides, P, Sam, C, Goldsworthy, H, Armijo, M & Tognatta, N 2013, *Evaluation of the i3 scale-up of Reading Recovery: Year one report, 2011-2012*, Consortium for Policy Research in Education, viewed 27 March 2015, http://cpre.org/sites/default/files/researchreport/1488_readingrecoveryreport.pdf
- May, H, Gray, A, Sirindes, P, Goldsworthy, H, Armijo, M, Sam, C, Gillespie, JN & Tognatta, N 2015, 'Year one results from the Randomized Evaluation of the i3 scale-up of Reading Recovery', *American Educational Research Journal*, advance online publication.

- Moats, L 2007, *Whole-language high jinks: How to tell when “scientifically-based instruction” isn’t*. Thomas Fordham Foundation.
- Reading Recovery Council of North America 2015, *Clay’s Observation Survey*, viewed 5 November 2015, <http://readingrecovery.org/reading-recovery/teaching-children/observation-survey>
- Reynolds, M & Wheldall, K 2007, ‘Reading Recovery 20 years down the track: Looking forward, looking back’ *International Journal of Disability, Development and Education*, vol. 54, no. 2, pp. 199-223.
- Rosenthal, R & Jacobson, L 1966, ‘Teachers’ expectancies: Determinants of Pupils’ IQ gains’ *Psychological Reports*, vol. 19, pp. 115-118.
- Rowe, K 2005, *Teaching Reading literature review: A review of the evidence-based research literature on approaches to the teaching of literacy, particularly those that are effective in assisting students with reading difficulties*. A report of the Committee for the National Inquiry into the Teaching of Literacy, Australian Government Department of Education, Science and Training, Canberra, ACT.
- Tunmer, WE & Chapman, JW 2003, ‘The Reading Recovery approach to preventative early intervention: As good as it gets?’, *Reading Psychology*, vol. 24, pp. 337-360.
- Tunmer, WE, Chapman, JW, Greaney, KT, Prochnow, JE. & Arrow, AW 2013, ‘Why the New Zealand National Literacy Strategy has failed and what can be done about it: Evidence from the Progress in International Reading Literacy Study (PIRLS) 2011 and Reading Recovery monitoring reports’ *Australian Journal of Learning Difficulties*, vol. 18, no. 2, pp. 139-180.
- U.S. Department of Health and Human Services 2000, *Report of the National Reading Panel: Teaching children to read –findings and determinations of the National Reading Panel by Topic Areas*, viewed 27 March 2015, <http://www.nichd.nih.gov/publications/pubs/nrp/pages/findings.aspx>
- What Works Clearinghouse 2008, *Intervention report: Reading Recovery*, United States Department of Education, Institute of Education Sciences.
- What Works Clearinghouse 2013, *What Works Clearinghouse Intervention Report: Reading Recovery*. United States Department of Education, Institute of Education Sciences.
- Willms, JD 2003, ‘Literacy proficiency of youth: evidence of converging socioeconomic gradients’, *International Journal of Educational Research*, vol. 39, pp. 247–252.

Appendix A. Sample Representativeness

A large proportion of the initial RR and non-RR cohorts identified in the 2011 Best Start assessment were excluded in determining the final samples included in the Year 1 and Year 3 regression analyses. There were several reasons for this, including: a failure to match student records (based on Student ID) from Best Start through the Literacy Continuum assessments and NAPLAN; the exclusion of non-RR students from RR schools; and issues with missing data, particularly on Literacy Continuum assessments conducted at the end of Kindergarten and Year 1. To examine whether the cohorts included in each regression analysis were representative of the total population of students, preliminary descriptive analyses were undertaken comparing the full cohort of students identified in the Best Start 2011 assessment separately with each sub-sample included in the end of Year 1 regression for Reading Texts and the Year 3 NAPLAN regression²⁴. For non-RR students, results summarised in Table A1 show significant differences in the demographic distribution across cohorts. Compared to non-RR students who participated in Best Start in 2011, those included in both regression sub-samples were significantly less likely to be: Aboriginal, NESB, and identified as ESL Phase 1. However, the students included in the regression sub-samples were each more likely to have parents with a Year 12 education and to have parents in management positions (all p-values < .001).

Table A1:

Demographic distribution of the non-RR students identified in the Best Start 2011 sample versus those included in the Year 1 and Year 3 regression samples

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

	Best Start (2011) (n=56,897)	Year 1 Reading Texts Regression (n = 3,408)	Year 3 NAPLAN Reading Regression (n = 6,499)
Gender (%)			
Female	49.5	49.1	49.3
Male	50.5	50.9	50.7
Aboriginal Status (%)			
Non-Aboriginal	92.2	93.8	92.9
Aboriginal	5.7	3.8	4.3
Unknown	2.1	2.4	2.8
NESB (%)			
No	71.1	75.7	76.8
Yes	28.9	24.3	23.2
ESL Phase (%)			
Phase 1	10.3	6.3	5.5
Phase 2	3.6	3.9	3.1
Phase 3	0.7	1.0	0.8
Other	85.4	88.8	90.7
Parent Education (%)^a			
Not stated	5.1	4.1	4.4
Year 9 or equivalent	5.6	3.6	3.8
Year 10 or equivalent	20.4	18.1	18.0
Year 11 or equivalent	6.9	5.6	5.8
Year 12 or equivalent	62.1	68.6	68.0

²⁴ Analysis of sample representativeness for the end of Year 1 regression sample focussed only on those included in the Reading Text regression. While separate regressions were undertaken for each Literacy Continuum aspect, the samples included in each did not differ significantly in their demographic or performance distributions.

Table A1:

Demographic distribution of the non-RR students identified in the Best Start 2011 sample versus those included in the Year 1 and Year 3 regression samples

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

	Best Start (2011) (n=56,897)	Year 1 Reading Texts Regression (n = 3,408)	Year 3 NAPLAN Reading Regression (n = 6,499)
Parent Occupation Group (%)^a			
Senior Management	14.5	18.8	19.1
Other Business Manager	16.6	19.2	19.5
Trades/Clerk/Sales	17.4	18.1	17.6
Machine Operators/Hospitality	11.8	10.3	10.5
Not in paid work for 12 months	26.7	23.5	22.6
Not stated	12.9	10.0	10.7

Note. Data for gender, Aboriginal status, NESB and ESL Phase are drawn from Best Start, while parent education and occupation are drawn from NAPLAN. ^aReported percentages exclude cases with missing parental education and occupation data (missing n = 8,200).

For RR students, results summarised in Table A2 showed largely similar demographic distributions for those included in Best Start and in each regression sub-sample. However, compared to all RR students who participated in Best Start, those included in the Year 1 Reading Texts regression were significantly more likely to be NESB and to be identified as ESL Phase 1, while they were significantly less likely to have parents in management positions. The only significant difference to emerge for RR students included in the Year 3 regression compared to the full Best Start cohort was that they were more likely to be classified as NESB (all p-values < .001).

Table A2:

Demographic distribution of the RR students identified in the Best Start 2011 sample versus those included in the Year 1 and Year 3 regression samples

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

	Best Start (2011) (n=8,638)	Year 1 Reading Texts Regression (n = 3,187)	Year 3 NAPLAN Reading Regression (n = 4,408)
Gender (%)			
Female	40.0	39.8	40.6
Male	60.0	60.2	59.4
Aboriginal Status (%)			
Non-Aboriginal	86.2	87.8	87.5
Aboriginal	11.4	10.1	10.1
Unknown	2.4	2.1	2.5
NESB (%)			
No	72.2	65.1	69.1
Yes	27.9	34.9	30.9
ESL Phase (%)			
Phase 1	12.7	15.6	14.0
Phase 2	2.2	2.4	2.3
Phase 3	0.4	0.5	0.5
Other	84.6	81.5	83.3
Parent Education (%)^a			
Not stated	7.5	6.8	6.7
Year 9 or equivalent	12.5	13.5	12.7
Year 10 or equivalent	30.3	29.5	30.0
Year 11 or equivalent	9.2	9.0	9.2
Year 12 or equivalent	40.5	41.2	41.4
Parent Occupation Group (%)^a			
Senior Management	5.5	4.8	5.3
Other Business Manager	9.9	8.6	9.3
Trades/Clerk/Sales	15.4	14.5	15.9
Machine Operators/Hospitality	13.9	14.1	14.4
Not in paid work for 12 months	34.1	36.9	34.8
Not stated	21.2	21.1	20.3

Note. Data for gender, Aboriginal status, NESB and ESL Phase are drawn from Best Start, while parent education and occupation are drawn from NAPLAN. ^aReported percentages exclude cases with missing parental education and occupation data (missing n = 781).

In relation to performance indicators, each regression sub-sample for Year 1 Reading Texts and Year 3 NAPLAN Reading scores was separately compared to the overall Best Start sample on the distribution of scores across each of the seven aspects contained in the 2011 Best Start assessment and on the mean Year 3 NAPLAN Reading scores. As summarised in Table A3, for non-RR students, there were significant differences between the full Best Start sample and those included in each regression sub-sample across all Best Start aspects (except Aspects of Writing) and on NAPLAN Year 3 scores (all p-values <.001). In each instance, results showed that the non-RR students included in each regression sub-sample were more likely to be assessed at higher levels in Best Start and to achieve higher mean NAPLAN Reading scores in Year 3 than the full cohort who participated in Best Start.

Table A3:

Distribution of Best Start assessment and Year 3 NAPLAN Reading scores for all non-RR students identified in the Best Start 2011 sample versus those included in the Year 1 and Year 3 regression samples

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery and Best Start data sourced from the Early Learning and Primary Education Directorate, DOE.

	Best Start (2011) (n=56,897)	Year 1 Reading Texts Regression (n = 3,408)	Year 3 NAPLAN Reading Regression (n = 6,499)
Reading Texts (%)			
Level 0	55.2	49.1	48.6
Level 1	42.0	48.0	48.1
Level 2	1.8	1.8	2.1
Level 3	1.0	1.1	1.3
Comprehension (%)			
Level 0	49.4	43.5	42.7
Level 1	41.5	45.9	46.2
Level 2	7.6	8.9	9.1
Level 3	1.4	1.7	2.0
Aspects of Writing (%)			
Level 0	78.7	77.7	77.6
Level 1	19.5	20.2	20.3
Level 2	1.6	2.0	1.9
Level 3	0.2	0.2	0.2
Aspects of Speaking (%)			
Level 0	31.2	24.4	24.2
Level 1	41.8	45.6	44.0
Level 2	25.0	28.0	29.1
Level 3	2.0	2.0	2.7
Phonics (%)			
Level 0	39.1	33.8	32.2
Level 1	41.0	43.5	43.6
Level 2	13.8	16.7	16.9
Level 3	6.1	6.0	7.4
Phonemic Awareness (%)			
Level 0	61.8	55.3	54.9
Level 1	30.6	36.4	36.1
Level 2	5.1	5.5	5.9
Level 3	2.5	2.7	3.2
Concepts about Print (%)			
Level 0	53.9	47.3	46.3
Level 1	40.9	46.2	46.7
Level 2	2.8	3.8	3.7
Level 3	2.4	2.7	3.3
NAPLAN Year 3 (2014):			
Mean Scaled Reading Score	431	439	439

Note. Reported means exclude students with missing scores. For the Best Start sample, less than 4%, on average, have missing Best Start scores and 17.4% have missing NAPLAN scores. For the Year 1 sample, less than 2%, on average, have missing Best Start Scores and 2.4% have missing NAPLAN scores. For the Year 3 sample, less than 1.5%, on average, have missing Best Start scores.

Results for the full cohort of RR students identified at Best Start and those sub-samples included in the Year 1 Reading Texts regression and the Year 3 NAPLAN regression are summarised in Table A4. The only significant differences to emerge for RR students was between the full Best Start cohort and the sub-sample included in the Year 1 regression on the distribution of Best Start levels for the aspects corresponding to Reading Texts, Comprehension, Aspects of Speaking and Phonics (all p-values <.05). In each instance, RR students included in the Year 1 regression sub-sample were more likely to be assessed at lower levels on these aspects compared to the full cohort of RR students who completed Best Start. There were no significant differences in mean NAPLAN Reading Scores in Year 3 for RR students included in the regression sub-samples and the full Best Start cohort.

Table A4:

Distribution of Best Start assessment and Year 3 NAPLAN Reading scores for all RR students identified in the Best Start 2011 sample versus those included in the Year 1 and Year 3 regression samples

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery and Best Start data sourced from the Early Learning and Primary Education Directorate, DOE.

	Best Start (2011) (n=8,638)	Year 1 Reading Texts Regression (n = 3,187)	Year 3 NAPLAN Reading Regression (n = 4,408)
Reading Texts (%)			
Level 0	77.5	81.3	78.8
Level 1	22.3	18.5	21.0
Level 2	0.2	0.2	0.2
Level 3	0.0	0.0	0.0
Comprehension (%)			
Level 0	71.1	73.8	72.1
Level 1	26.5	24.1	25.5
Level 2	2.3	2.0	2.2
Level 3	0.2	0.2	0.2
Aspects of Writing (%)			
Level 0	90.6	91.5	90.8
Level 1	9.2	8.3	9.0
Level 2	0.2	0.2	0.2
Level 3	0.0	0.0	0.0
Aspects of Speaking (%)			
Level 0	53.8	57.9	54.3
Level 1	35.7	33.2	35.6
Level 2	10.1	8.7	9.7
Level 3	0.4	0.3	0.5
Phonics (%)			
Level 0	73.4	76.5	74.6
Level 1	24.7	21.9	23.5
Level 2	1.7	1.5	1.7
Level 3	0.3	0.1	0.2
Phonemic Awareness (%)			
Level 0	86.5	88.0	86.6
Level 1	13.0	11.5	12.9
Level 2	0.5	0.4	0.5
Level 3	0.0	0.0	0.0
Concepts about Print (%)			
Level 0	80.3	82.5	80.6
Level 1	19.2	17.0	18.9
Level 2	0.4	0.4	0.3
Level 3	0.1	0.2	0.2
NAPLAN Year 3 (2014):			
Scaled Reading Score	330	330	331

Note. Reported means exclude students with missing scores. For the Best Start sample, less than 5%, on average, have missing Best Start scores and 14.1% have missing NAPLAN scores. For the Year 1 sample, less than 2%, on average, have missing Best Start Scores and 5% have missing NAPLAN scores. For the Year 3 sample, less than 1.5%, on average, have missing Best Start scores.

Appendix B. Year 1 Regression Technical Information

Scores on the Literacy Continuum collected at the end of Year 1 (Y_{ij}) were treated as ordinal variables. In ordered logistic regression models, an underlying score (Y_{ij}^*) is estimated as a linear function of the independent variables, random effects and a set of threshold points (see equation (1)). The probability of observing level k corresponds to the probability that the estimated linear function, plus residual errors, is within the range of threshold points estimated for the Literacy Continuum level at the end of Year 1 (see equation (2)).

$$(1) Y_{ij}^* = \mathbf{bX} + \mathbf{cW} + u_{oj} + e_{ij}$$

$$(2) P(Y_{ij} = k) = P(a_{k-1} < Y_{ij}^* < a_k) = \frac{1}{1 + \exp(-a_k + \mathbf{bX} + \mathbf{cW} + u_{oj})} - \frac{1}{1 + \exp(-a_{k-1} + \mathbf{bX} + \mathbf{cW} + u_{oj})}$$

where i represents student i ; j represents school j ; k represents the category for the level of the Literacy Continuum at the end of Year 1 ($k=0,1,\dots,M$); M is the last category; a_k represents the cut of points between the $k-1$ th and the k th category ($k=1,\dots,M-1$); a_0 is taken as negative infinity; a_M is taken as positive infinity; and \mathbf{bX} and \mathbf{cW} are linear functions of student-level variables and school-level variables respectively as specified below:

(3) Level 1 (student):

$$\begin{aligned} \mathbf{bX} = & b_1 X_{1_1ij} + b_2 X_{1_2ij} + b_3 X_{1_3ij} + b_4 X_{2ij} + b_5 X_{1_1ij} X_{2ij} + b_6 X_{1_2ij} X_{2ij} + b_7 X_{1_3ij} X_{2ij} + b_8 X_{3_1ij} + b_9 \\ & X_{3_2ij} + b_{10} X_{3_1ij} X_{2ij} + b_{11} X_{3_2ij} X_{2ij} + b_{12} X_{4_1ij} + b_{13} X_{5_1ij} + b_{14} X_{5_2ij} + b_{15} X_{6ij} + b_{16} X_{7_1ij} + b_{17} X_{8_1ij} + \\ & b_{18} X_{8_2ij} + b_{19} X_{8_3ij} + b_{20} X_{8_4ij} + b_{21} X_{9_1ij} + b_{22} X_{9_2ij} + b_{23} X_{9_3ij} + b_{24} X_{9_4ij} + b_{25} X_{10_1ij} + b_{26} X_{10_2ij} + \\ & b_{27} X_{10_3ij} + b_{28} X_{10_4ij} + b_{29} X_{11_1ij} + b_{30} X_{11_2ij} \end{aligned}$$

(4) Level 2 (school):

$$\begin{aligned} \mathbf{cX} = & c_1 W_{1ij} + c_2 W_{2_1ij} + c_3 W_{2_2ij} + c_4 W_{3ij} + c_5 W_{4_1ij} + c_6 W_{4_2ij} + c_7 W_{4_3ij} + c_8 W_{4_4ij} + c_9 W_{4_5ij} + c_{10} W_{4_6ij} \\ & + c_{11} W_{4_7ij} + c_{12} W_{4_8ij} + c_{13} W_{4_9ij} \end{aligned}$$

u_{oj} represents the unobserved school-specific effect of school j ; e_{ij} represents the residual of student i in school j ; $b_1, \dots, b_{30}, c_1, \dots, c_{13}$ are the regression coefficients to be estimated. The set of student-level and school-level variables are listed below:

$$\begin{aligned} X_{1_1ij} &= 1 \text{ if Literacy Continuum level at Term 4 Kindergarten} = 2 \text{ and } 0 \text{ otherwise;} \\ X_{1_2ij} &= 1 \text{ if Literacy Continuum level at Term 4 Kindergarten} = 3 \text{ and } 0 \text{ otherwise;} \\ X_{1_3ij} &= 1 \text{ if Literacy Continuum level at Term 4 Kindergarten} = 4 \text{ or above and } 0 \text{ otherwise;} \\ X_{2ij} &= 1 \text{ if student } i \text{ has received RR intervention and } 0 \text{ otherwise;} \end{aligned}$$

X_{3_1ij}	=	1 if Best Start Literacy assessment level at Term 1 Kindergarten = k and 0 otherwise ($k=1$);
X_{3_2ij}	=	1 if Best Start Literacy assessment level at Term 1 Kindergarten = 2 or 3 and 0 otherwise;
X_{4ij}	=	1 if student i is a male and 0 otherwise;
X_{5_1ij}	=	1 if student i is Aboriginal and 0 otherwise;
X_{5_2ij}	=	1 if student i 's Aboriginal status is unknown and 0 otherwise;
X_{6ij}	=	1 if student i is born in Australia or external territories and 0 otherwise;
X_{7ij}	=	1 if student i has non-English speaking background and 0 otherwise;
X_{8_1ij}	=	1 if student i is at ESL Phase 2 at Best Start (Term 1 Kindergarten) and 0 otherwise;
X_{8_2ij}	=	1 if student i is at ESL Phase 3 at Best Start (Term 1 Kindergarten) and 0 otherwise;
X_{8_3ij}	=	1 if student i 's ESL Phase at Best Start (Term 1 Kindergarten) is recorded as to be assessed/unknown/not provided and 0 otherwise;
X_{8_4ij}	=	1 if student i 's ESL Phase at Best Start (Term 1 Kindergarten) is recorded as not required and 0 otherwise;
X_{9_1ij}	=	1 if student i 's parent 1 has no non-school education and 0 otherwise;
X_{9_2ij}	=	1 if student i 's parent 1 has completed Certificate I to IV (including trade certificate) and 0 otherwise;
X_{9_3ij}	=	1 if student i 's parent 1 has completed Advanced diploma/Diploma and 0 otherwise;
X_{9_4ij}	=	1 if student i 's parent 1 has completed Bachelor degree or above and 0 otherwise;
X_{10_1ij}	=	1 if student i 's parent 1 occupation group is recorded as "Other business manager, arts/media/sportspersons and associate professionals" and 0 otherwise;
X_{10_2ij}	=	1 if student i 's parent 1 occupation group is recorded as "Tradesmen/women, clerks and skilled office, sales and service staff" and 0 otherwise;
X_{10_3ij}	=	1 if student i 's parent 1 occupation group is recorded as "Machine operators, hospitality staff, assistants, labourers and related workers" and 0 otherwise;
X_{10_4ij}	=	1 if student i 's parent 1 has not been in paid work for more than 12 months and 0 otherwise;
X_{11_1ij}	=	1 if student i 's parent 1 is speaking a language other than English at home and 0 otherwise;
X_{11_2ij}	=	1 if student i 's parent 1 speaking language at home is listed as unknown and 0 otherwise;
W_{1ij}	=	1 if school j participates in the Priority Funding Schools Program (PSFP) and 0 otherwise;
W_{2_1ij}	=	1 if school j location is "Provincial" and 0 otherwise;
W_{2_2ij}	=	1 if school j location is "Remote" or "Very remote" and 0 otherwise;
W_{3ij}	=	Family Occupation and Education Index (FOEI) of school j in 2013
W_{4_1ij}	=	1 if school j is in "Northern Sydney" region and 0 otherwise;
W_{4_2ij}	=	1 if school j is in "South Western Sydney" region and 0 otherwise;
W_{4_3ij}	=	1 if school j is in "Western Sydney" region and 0 otherwise;
W_{4_4ij}	=	1 if school j is in "Hunter/Central Coast" region and 0 otherwise;
W_{4_5ij}	=	1 if school j is in "North Coast" region and 0 otherwise;
W_{4_6ij}	=	1 if school j is in "New England" region and 0 otherwise;
W_{4_7ij}	=	1 if school j is in "Riverina" region and 0 otherwise;
W_{4_8ij}	=	1 if school j is in "Illawarra and South East" region and 0 otherwise; and
W_{4_9ij}	=	1 if school j is in "Western NSW" region and 0 otherwise.

In the model specified in equations (1) and (2), the residual errors (e_{ij}) are assumed to follow a logistic distribution with zero mean and variance $\pi^2/3$ and are independent of the random school-specific effects (u_{0j}). The school-specific random effects (u_{0j}) are assumed to be normally distributed with zero mean and constant variance (σ_{u0^2}). The variances of random effects (σ_{u0^2}) are also known as the between-school variance.

To predict the Literacy Continuum level for a student, an underlying score (Y_{ij}^*) was calculated based on equation (1). Then the level was determined based on where the underlying score fell between the threshold points. For example, if $a_1 < Y_{ij}^* < a_2$, then the predicted level is 2 for a student. The ordered logistic regression model is sometimes written in the following form (see equation (5)) to represent the odds of getting a level greater than k .

$$(5) \text{ Odds } (k) = \frac{P(Y_{ij} > k)}{P(Y_{ij} \leq k)} = \frac{\mathbf{bX} + \mathbf{cW} + u_{0j} + e_{ij}}{P(Y_{ij} \leq k)} \quad \text{where } k = 0, 1, \dots, M.$$

Note that $Odds(0) = Odds(1) = \dots = Odds(M)$.

That is, the model assumes that the effect of each independent variable is the same for all categories of the dependent variable. Therefore the ordered logistic regression model is also known as the proportional odds model. The Brant test was used to test for the proportional odds assumption for the ordered logistic model without random effects (Long and Freese, 2006). It was found that a few independent variables violated this assumption. While some literature suggests that the violation of this assumption does not impact the practical implication of the coefficients and requires no further action, other literature suggests extending the model to relax the proportional odds assumption. For example, Anderson (1984) proposed the stereotype model to relax the proportional odds assumption by adding extra parameters in the model. Taking this approach, a stereotype ordered logistic regression model was fitted to the current data and the regression coefficients were compared with those derived from the ordered logistic model. The results showed that the coefficients and associated significance levels did not differ markedly from the ordered logistic regression model. Therefore, the ordered logistic model with random effects was adopted for analysis.

Appendix C. Propensity Score Matching

The potential presence of unmeasured confounding factors raises concerns about the presence of omitted variable bias in estimating the effect of RR on student outcomes. To correct the bias, Rosenbaum and Rubin (1983) proposed propensity score matching (PSM), which aims to approximate the conditions of randomisation by matching pairs of individuals who are equally likely to receive RR intervention on the basis of their observed characteristics. The first step of PSM involves fitting a logistic regression model to predict the likelihood of students receiving RR using all observable covariates known or expected to influence selection into RR and/or known or expected to be related to the outcome variable (i.e. Literacy Continuum aspect level at the end of Year 1 or NAPLAN Reading score at the end of Year 3). This model included covariates for all student-level and school-level variables used in the random effects model (as listed above). The predicted probability yielded from the model is known as the propensity score.

The second step is to match each RR student with a non-RR student using the smallest difference in the propensity scores. In the current analysis, a one-to-one nearest neighbour matching with replacement was applied so that a non-RR student could be considered a best match for more than one RR student. The maximum difference in propensity score (caliper) was set to be 0.0001 for a pair of students to be matched. After matching, all of the covariates are examined to test whether they are “balanced” (equal in mean) amongst the matched pairs of students. Three measures, (1) t-test for equality of means; (2) standardised bias (SB); and (3) likelihood ratio test of the joint insignificance of all independent variables, were used to assess the balancing of each variable. T-tests are used to test if the mean of the covariate differs between non-RR students and RR students before and after matching. The p-value is expected to be greater than 0.05 after matching when the covariate is balanced between the two groups. The second measure, SB, is defined by Rosenbaum and Rubin (1985) and is computed as:

$$(8) \text{ SB} = \frac{100 (\bar{x}_1 - \bar{x}_2)}{\sqrt{(s_1^2 - s_2^2)/2}}$$

where \bar{x}_1 and \bar{x}_2 are the means of the covariates for the RR students and non-RR students respectively; s_1^2 and s_2^2 represent the variances of covariates. The absolute value of SB is expected to be less than 20 if the covariate is balanced. The likelihood ratio test was used to test if all independent variables in the logistic model were insignificant. The p-value is expected to be greater than 0.05 after matching when the covariate is balanced between the two groups.

Once the covariates are balanced on all observable measures, the last step of PSM is to compare the matched pairs to determine whether the RR effect is significant, or in other words, whether RR students performed better than non-RR students on Literacy Continuum aspects at Term 4 Year 1 and on the NAPLAN Reading domain in Year 3. A simple ordered logistic model controlling for the nature of the matched pairs was fitted for the Literacy Continuum outcomes at the end of Year 1 on the following independent variables: (1) Literacy Continuum aspect level at Term 4 Kindergarten; (2) whether the students received RR or not; (3) the interaction between the first two variables; and (4) the EAL/D phase at Term 4 Year 1. For each level of the Literacy Continuum aspects at Term 4 Kindergarten, an odds ratio significantly greater than one (p-value < .05) indicates that RR students are more likely to progress to higher levels on the Literacy Continuum at the end of Year 1 relative to non-RR students. Odds ratios significantly less than one indicate that non-RR students are more likely to progress on the Continuum aspects relative to their RR counterparts.

Similarly, a simple linear regression model was fitted for NAPLAN Reading scores in Year 3 controlling for the same four independent variables as listed above and the matched pair nature of the sample. The RR effect is computed as the difference in the predicted NAPLAN Reading score between RR students and non-RR students. A significantly positive difference (p -value $< .05$) indicates that the RR students were more likely to achieve higher NAPLAN Reading scores relative to non-RR students. Significantly negative differences indicate that non-RR students were more likely to achieve higher NAPLAN Reading scores relative to their RR counterparts.

It is important to note that PSM is not a substitute for a randomised controlled trial. Comparability between RR students and non-RR students is only achieved on measured covariates and the possibility of omitted variable bias still exists.

Appendix D. Year 3 Regression Technical Information

A random effects linear regression model was used to examine the effect of RR on the raw NAPLAN Reading score. The model is specified as follows for the NAPLAN Reading score (Y_{ij}):

$$(6) Y_{ij} = \mathbf{bX} + \mathbf{cW} + u_{oj} + e_{ij}$$

The linear function \mathbf{bX} is defined slightly differently compared to equation (3) with the addition of two extra categorical independent variables: the Comprehension level at Term 4 Kindergarten and the Best Start Comprehension level assessed at Term 1 Kindergarten. The linear function \mathbf{cW} is the same as equation (4) in Appendix B. The model predicts the NAPLAN Reading score of student i in school j (Y_{ij}) in Year 3. This is also known as a random intercepts model because only the intercepts (u_{oj}) are random at the school level. The model assumptions for the random effects model include: (1) normality of errors; (2) homoscedasticity of errors; and (3) independence of observations. The first assumption states that the residual errors (e_{ij}) and the school-specific random effects (u_{oj}) are normally distributed with zero mean and constant variance σ_e^2 and σ_{u0}^2 respectively. The variances of the two errors are known as the within-school variance (σ_e^2) and between-school variance (σ_{u0}^2). The second assumption requires the variances of the errors (σ_e^2 and σ_{u0}^2) to be constant. The third assumption specifies that the students are random samples from a larger population and that the student's NAPLAN Reading scores are independent of each other. All assumptions were met in the current analysis.

Appendix E. Year 1 Supplementary Results

Table E1:

Odds ratios of random effects ordered logistic regression models across the seven aspects of Literacy Continuum for non-RR and RR students (discontinued and referred) at the end of Year 1.

Sources: Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

Variables	Category	Reading Texts	Comprehension	Aspects of Writing	Aspects of Speaking	Phonics	Phonemic awareness	Concepts about print
Corresponding Lit. Cont. aspect level T4K	Level 1 or below	–	–	–	–	–	–	–
	Level 2	5.326*	9.121*	8.814*	4.795*	3.538*	5.606*	6.594*
	Level 3	39.591*	36.529*	45.534*	15.395*	23.765*	20.341*	28.093*
	Level 4+	328.19*	190.99*	299.50*	54.290*	132.57*	105.71*	649.31*
RR	No	–	–	–	–	–	–	–
	Yes	2.666*	1.295	0.975	0.784	0.962	0.853	1.109
Interaction: Lit. Cont. level at T4K x RR	Level 2#RR	0.563*	0.409*	0.407*	0.652*	1.045	0.605*	0.478*
	Level 3#RR	0.157*	0.230*	0.228*	0.562*	0.294*	0.343*	0.293*
	Level 4+#RR	0.034*	0.096*	0.115*	0.438*	0.148*	0.261*	0.049*
Best Start (BS) level (T1K)	Level 0	–	–	–	–	–	–	–
	Level 1	1.807*	1.399*	1.661*	1.756*	1.552*	1.656*	1.305
	Level 2 or 3	6.597*	1.670*	3.220*	2.695*	2.851*	2.435*	1.085
Interaction: BS level (T1K) x RR	Level 1#RR	0.690*	0.968	0.786	0.913	0.872	0.752	1.240
	Level 2 or 3#RR	0.053*	1.157	0.654	0.624*	0.462*	0.482	1.508
Gender	Female	–	–	–	–	–	–	–
	Male	0.849*	0.881*	0.701*	0.745*	0.845*	0.798*	0.794*
Aboriginal status	Non-Aboriginal	–	–	–	–	–	–	–
	Aboriginal	0.735*	0.911	0.803*	0.778*	0.756*	0.675*	0.656*
	Unknown	1.027	0.928	0.826	0.797	0.808	0.864	1.116
Country of birth	Outside Australia	–	–	–	–	–	–	–
	Australia and external territories	0.709*	0.654*	0.752*	0.733*	0.886	0.767*	0.753
Non-English speaking background	No	–	–	–	–	–	–	–
	Yes	0.949	1.168	1.038	0.861	0.995	1.060	0.879
ESL Phase at BS (T1K)	Phase 1	–	–	–	–	–	–	–
	Phase 2	0.832	1.205	1.111	1.443	0.942	1.254	1.632*
	Phase 3	0.805	1.076	1.149	0.798	1.143	1.271	1.034
	To be assessed/Unknown/Not provided	1.027	1.322*	1.166	1.231	0.951	1.117	1.080
	Not required	1.107	1.312	1.220	1.256	1.259	1.232	1.379
Parent 1 qualification	Not Stated	–	–	–	–	–	–	–
	No Non-school education	1.034	1.014	1.162	1.023	1.039	1.072	1.212
	Certificate I to IV (including trade certificate)	1.166	1.130	1.209*	1.105	1.027	1.094	1.198
	Advanced diploma/Diploma	1.283*	1.175	1.277*	1.253*	1.149	1.193	1.292
	Bachelor degree or above	1.752*	1.620*	1.722*	1.738*	1.575*	1.508*	1.466*
Parent 1 occupation group	Senior management	–	–	–	–	–	–	–
	Other business manager	1.172	0.975	0.985	0.903	1.003	1.184	1.118
	Tradesmen/women, Clerks, Sales	1.032	0.873	0.798*	0.924	0.935	0.856	0.829
	Machine operators, Hospitality staff	0.917	0.853	0.821	0.764*	0.963	0.937	0.790
	Not in Paid work in last 12 months	1.050	0.930	0.854	0.818*	0.950	0.849	0.863
	Not stated	1.015	1.012	0.844	0.821	1.016	0.992	0.875

Table E1:

Odds ratios of random effects ordered logistic regression models across the seven aspects of Literacy Continuum for non-RR and RR students (discontinued and referred) at the end of Year 1.

Sources: Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

Variables	Category	Reading Texts	Compre-hension	Aspects of Writing	Aspects of Speaking	Phonics	Phonemic awareness	Concepts about print
Parent 1 speaking language other than English at home	No	–	–	–	–	–	–	–
	Yes	1.265	0.989	1.141	1.168	1.244	1.221	1.017
	Unknown	1.256	1.701	1.369	0.982	1.421	1.476	1.426
PSFP at T4K	No	–	–	–	–	–	–	–
	Yes	1.130	0.678*	0.729*	0.711*	0.760*	0.829	0.743
School location at T4 Y1	Metropolitan	–	–	–	–	–	–	–
	Provincial	0.858	0.695	0.581*	0.792	0.690	0.492*	0.551
	Remote/very remote	2.298	1.229	0.363	1.031	1.095	5.551*	0.794
FOEI at T4 Y1		0.996	1.000	1.000	1.000	1.000	1.001	1.002
School region at T4 Y1	Sydney	–	–	–	–	–	–	–
	Northern Sydney	2.115	1.568	1.097	1.941	2.153	1.700	0.801
	South Western Sydney	1.221	1.264	0.966	1.239	1.220	1.248	1.567
	Western Sydney	1.495	1.260	1.130	1.800*	1.359	1.728*	1.785
	Hunter/Central Coast	1.243	1.100	0.881	0.939	1.336	1.346	1.168
	North Coast	1.771	1.964	1.867	1.81	2.091*	3.683*	3.497*
	New England	1.226	1.429	1.479	1.629	2.081	2.672*	1.969
	Riverina	2.218	2.513*	2.554*	1.991	2.762*	3.393*	5.066*
	Illawarra and South East	2.008*	2.291*	2.054*	1.888*	2.406*	3.473*	4.453*
Western NSW	1.732	1.263	1.889	1.302	1.957	2.554	1.853	
Model N	n	6,595	6,536	6,484	6,570	6,613	6,543	6,573

Note. * indicates that the odds ratio is statistically significant at the 0.05 level.

Table E2:

Estimated effect of RR on the Literacy Continuum level achieved at the end of Year 1 using PSM approach

Sources: Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

Lit. Cont. Aspect	Lit. Cont. Aspect Level (T4K)	Odds Ratio	Mean [SB]	No. of matched pairs
Reading Texts	Level 1 or below	1.94*	3.2	759
	Level 2	1.17		
	Level 3	0.40*		
	Level 4 or above	0.16*		
Comprehension	Level 1 or below	0.74	3.0	789
	Level 2	0.65*		
	Level 3	0.39*		
	Level 4 or above	0.13*		
Aspects of Writing	Level 1 or below	0.96	3.3	804
	Level 2	0.55*		
	Level 3	0.32*		
	Level 4 or above	0.28		
Aspects of Speaking	Level 1 or below	0.68	2.4	837
	Level 2	0.54*		
	Level 3	0.46*		
	Level 4 or above	0.20*		
Phonics	Level 1 or below	0.90	2.2	752
	Level 2	1.05		
	Level 3	0.38*		
	Level 4 or above	0.19*		
Phonemic Awareness	Level 1 or below	0.85	3.3	761
	Level 2	0.61*		
	Level 3	0.36*		
	Level 4 or above	0.15*		
Concepts about Print	Level 1 or below	1.68	2.5	838
	Level 2	0.74		
	Level 3	0.69*		
	Level 4	<0.01*		

Note. For each Literacy Continuum aspect, the odds ratio is estimated for each aspect level in Term 4 Kindergarten. It represents the odds ratio of achieving a higher level in Literacy Continuum at the end of Year 1 for RR students relative to non-RR students. *Indicates that the odds ratio is significantly different from one at the .05 significance level. Results coloured in **cyan** show that RR students had higher odds of progressing on the Continuum relative to non-RR students; results coloured in **pink** show that non-RR students had higher odds of progressing on the Continuum relative to RR students.

Table E3:

Estimated effect of RR on the Literacy Continuum level achieved at the end of Year 1, including only discontinued students in the RR cohort

Sources: Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

Lit. Cont. Aspect	Lit. Cont. Aspect Level (T4K)	Odds Ratio	Lit. Cont. Aspect	Lit. Cont. Aspect Level (T4K)	Odds Ratio
Reading Texts	Level 1 or below	7.79*	Phonics	Level 1 or below	2.64*
	Level 2	2.14*		Level 2	1.41
	Level 3	0.43*		Level 3	0.32*
	Level 4 or above	0.08*		Level 4 or above	0.14*
Comprehension	Level 1 or below	2.49*	Phonemic Awareness	Level 1 or below	1.75*
	Level 2	0.64*		Level 2	0.61*
	Level 3	0.33*		Level 3	0.29*
	Level 4 or above	0.12*		Level 4 or above	0.22*
Aspects of Writing	Level 1 or below	2.02*	Concepts about Print	Level 1 or below	3.61*
	Level 2	0.47*		Level 2	0.73
	Level 3	0.23*		Level 3	0.39*
	Level 4 or above	0.10*		Level 4	0.06*
Aspects of Speaking	Level 1 or below	1.31			
	Level 2	0.61*			
	Level 3	0.49*			
	Level 4 or above	0.35*			

Note. For each Literacy continuum aspect, odds ratio is estimated for each aspect level in Term 4 Kindergarten. It represents the odds ratio of achieving a higher level in Literacy Continuum at the end of Year 1 for RR students relative to non-RR students. *Indicates that the odds ratio is significantly different from one at the .05 significance level. Results coloured in cyan show that RR students had higher odds of progressing on the Continuum relative to non-RR students; results coloured in pink show that non-RR students had higher odds of progressing on the Continuum relative to RR students.

Table E4:

Estimated effect of RR on the Literacy Continuum level achieved at the end of Year 1, excluding students who moved schools between the end of Kindergarten and the end of Year 1

Sources: Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

Lit. Cont. Aspect	Lit. Cont. Aspect Level (T4K)	Odds Ratio	Lit. Cont. Aspect	Lit. Cont. Aspect Level (T4K)	Odds Ratio
Reading Texts	Level 1 or below	2.79*	Phonics	Level 1 or below	0.91
	Level 2	1.62*		Level 2	0.99
	Level 3	0.45*		Level 3	0.28*
	Level 4 or above	0.09*		Level 4 or above	0.14*
Comprehension	Level 1 or below	1.33	Phonemic Awareness	Level 1 or below	0.79
	Level 2	0.52*		Level 2	0.50*
	Level 3	0.31*		Level 3	0.27*
	Level 4 or above	0.11*		Level 4 or above	0.21*
Aspects of Writing	Level 1 or below	1.07	Concepts about Print	Level 1 or below	0.98
	Level 2	0.43*		Level 2	0.50*
	Level 3	0.22*		Level 3	0.29*
	Level 4 or above	0.12*		Level 4	0.07*
Aspects of Speaking	Level 1 or below	0.77			
	Level 2	0.53*			
	Level 3	0.46*			
	Level 4 or above	0.34*			

Note. For each Literacy continuum aspect, odds ratio is estimated for each aspect level in Term 4 Kindergarten. It represents the odds ratio of achieving a higher level in Literacy Continuum at the end of Year 1 for RR students relative to non-RR students. *Indicates that the odds ratio is significantly different from one at the .05 significance level. Results coloured in **cyan** show that RR students had higher odds of progressing on the Continuum relative to non-RR students; results coloured in **pink** show that non-RR students had higher odds of progressing on the Continuum relative to RR students.

Appendix F. Year 3 Supplementary Results

Table F1:

Parameter estimates of random effects linear regression models on Year 3 NAPLAN Reading scores for non-RR students and RR students (discontinued and referred)

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

Variables	Category	Parameter estimates
Reading Texts level at T4K	Level 1 or below	–
	Level 2	11.864*
	Level 3	44.925*
	Level 4 or above	82.333*
RR	No	–
	Yes	-25.158*
Interaction: Reading Texts level at T4K x RR	Level 2#RR	0.284
	Level 3#RR	-27.966*
	Level 4+#RR	-61.512*
BS Reading Texts level (T1K)	Level 0	–
	Level 1	11.756*
	Level 2 or 3	37.878*
Comprehension level at T4K	Level 1 or below	–
	Level 2	11.766*
	Level 3	24.822*
	Level 4+	41.208*
BS Comprehension level (T1K)	Level 0	–
	Level 1	11.493*
	Level 2 or 3	34.788*
Gender	Female	–
	Male	-3.129*
Aboriginal status	Non-Aboriginal	–
	Aboriginal	-16.010*
	Unknown	9.281*
Country of birth	Outside Australia	–
	Australia and external territories	-6.849*
Non-English speaking background	No	–
	Yes	5.389
ESL Phase at BS (T1K)	Phase 1	–
	Phase 2	-9.172*
	Phase 3	-18.097*
	To be assessed/ Unknown/Not provided	2.869
	Not required	-4.876
Parent 1 qualification	Not stated	–
	No non-school education	1.125
	Certificate I to IV (including trade certificate)	2.227
	Advanced diploma/Diploma	9.046*
	Bachelor degree or above	20.915*

Table F1:

Parameter estimates of random effects linear regression models on Year 3 NAPLAN Reading scores for non-RR students and RR students (discontinued and referred)

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

Variables	Category	Parameter estimates
Parent 1 occupation group	Senior management	–
	Other business manager	-3.133
	Tradesmen/women, Clerks, Sales	-5.237
	Machine operators, Hospitality staff	-4.973
	Not in Paid work in last 12 months	-2.948
	Not Stated	-4.731
Parent 1 speaking language other than English at home	No	–
	Yes	4.377
	Unknown	-1.185
PSFP at T4K	No	–
	Yes	-2.917
School location at T4 Y1	Metropolitan	–
	Provincial	-1.214
	Remote	-4.203
	Very remote	40.547
FOEI at T4 Y1	–	-0.299*
School region at T4 Y1	Sydney	–
	Northern Sydney	1.399
	South Western Sydney	2.758
	Western Sydney	-1.535
	Hunter/Central Coast	-5.554
	North Coast	-3.189
	New England	-6.877
	Riverina	-7.591
	Illawarra and South East	-10.514*
Western NSW	2.652	
Constant	–	367.423*
Model N	n	10,907

Note. *Indicates that the coefficient is significantly different from zero at 0.05 significance level.

Table F2:

Estimated effect of RR on NAPLAN Reading outcomes in Year 3 using PSM approach

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

Reading Texts level at T4 K	RR effect on NAPLAN Reading score	p-value	Mean SB	No. of matched pairs
Level 1 or below	-33.2	<.001	1.8	1425
Level 2	-28.6	<.001		
Level 3	-58.0	<.001		
Level 4 or above	-88.1	<.001		

Note. Results coloured in **pink** show that RR students achieved lower NAPLAN Reading scores compared to non-RR students.

Table F3:


Estimated effect of RR on NAPLAN Reading outcomes in Year 3, including only discontinued students in the RR cohort

Sources: NAPLAN data extracted from the Statistics Unit, CESE. Reading Recovery, Best Start and Literacy Continuum data sourced from the Early Learning and Primary Education Directorate, DOE.

Reading Texts level at T4 K	RR effect on NAPLAN Reading score	p-value
Level 1 or below	-12.0	.034
Level 2	-20.9	<.001
Level 3	-52.0	<.001
Level 4 or above	-87.1	<.001

Note. Results coloured in pink show that RR students achieved lower NAPLAN Reading scores compared to non-RR students.

Centre for Education Statistics and Evaluation
GPO Box 33
Sydney NSW 2001
Australia

 02 9561 1211

 cese@det.nsw.edu.au

 www.cese.nsw.gov.au

NSW Department of Education
November 2015

