Abstract

Reading Recovery is an intensive one-to-one reading intervention programme designed for five- to six-year-old children who are the lowest literacy achievers after one year of formal tuition. Reading Recovery has been shown to have impressive effects in the short-term, particularly on those measures tailored to, and designed for, the programme. However, less is known about the programme's long-term effectiveness. The present study followed up at the end of Year 4: 120 comparison children, 73 children who had received Reading Recovery three years earlier, and 48 children in Reading Recovery schools who had not received Reading Recovery. We found that the children who had received Reading Recovery achieved an average National Curriculum level of 3b in reading which indicates being on track for Level 4 at the end of Key Stage 2. The comparison children were on average at Level 2a in reading, significantly lower than the Reading Recovery children. Reading Recovery children were also significantly less likely than comparison children to be identified as having a special educational need at the end of Year 3. These findings indicate that effects of the Reading Recovery programme are still apparent three years post-intervention.

Introduction

One of the key tasks of schooling is to ensure that children become confident readers and writers, able to access the curriculum and to be prepared for the myriad of demands on their literacy skills in adult life. It is now widely accepted that children with reading difficulties should be offered early intervention, and this is supported by the evidence of its short-term effectiveness (e.g., Wasik & Slavin, 1993; Torgesen, 2000; National Reading Panel, 2000). Early intervention offers an opportunity to prevent a widening gap

between poor readers and their peers as they move through school (Stanovich, 1986; Chall, 1983). Without action, poor readers read less than their peers (Allington, 1984; Biemiller, 1977-78; Clay, 1967; Juel, 1988), which in turn holds back their language development, their general knowledge, and even their IQ (Stanovich, 1986).

The aim of Every Child a Reader (ECAR), a government strategy in the UK, is to target those with reading difficulties (mostly living in poverty) and make sure that they are as literate as other six-year-olds. It is a school wide early literacy strategy involving a layered approach to intervention, with Reading Recovery at the core. ECAR schools are supported to provide high quality classroom instruction,, small group intervention with teaching assistants and finally, for the weakest readers, Reading Recovery. Designed by Marie Clay in New Zealand (Clay, 1985), Reading Recovery is an intensive, one-to-one reading intervention programme designed for five- to six-year-old children who are the lowest literacy achievers after one year of formal tuition. The programme involves daily, thirty minute lessons with a trained Reading Recovery teacher, over a period of up to twenty weeks, in which time the child is either discontinued; they reach the average reading level of their class, or referred; they are identified as potentially requiring further ongoing assessment. While the precise content of each lesson is tailored to the child and builds on what the child knows and needs to learn, sessions typically focus on developing the child's understanding of sounds, letters, words, and text, by getting them to draw upon their understanding of concepts about print, their phonological awareness, their understanding of meaning, and their knowledge of syntax. However, Reading Recovery is not only a remedial programme but a system for supporting children and teachers which ensures programme fidelity. This involves a one-year training for teachers, a tutor

who oversees a group of teachers and their practice and an assessment regime which underpins programme monitoring (Hurry, 1996). Evidence from the literature (e.g., Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994) has demonstrated strong short-term effects of Reading Recovery on reading and writing (although see Chapman, Tunmer, & Prochnow, 2001). However, there is little information on the durability of the gains observed as a result of the early interventions. There is also a lack of literature which has investigated the impact of Reading Recovery on children's Special Educational Needs status over time. In this paper, we report the effects of Reading Recovery on children's literacy progress and Special Educational Needs status three years after intervention.

The long-term effectiveness of Reading Recovery

In contrast to the extensive literature on the short-term effectiveness of Reading Recovery, many scholars (e.g., Hurry & Sylva, 2007) have pointed out that only a small evidence base exists for the programme's long-term effectiveness. This is problematic because information on the 'retention' of gains resulting from an early intervention (e.g., from the Reading Recovery programme) is paramount to any education system that might consider implementing that programme (Schwartz, Hobsbaum, Briggs, & Scull, 2009). The published literature on the long-term effectiveness of Reading Recovery is relatively sparse, but will now be considered.

Schmitt and Gregory (2005) compared the reading performance of US children who had been randomly selected for Reading Recovery in the First Grade with a cohort group of children who did not receive the intervention. It was found that the majority of children who had received Reading Recovery maintained their gains one, two, and three years later and were performing at a level equivalent with their same-aged peers on

measures of oral text reading. Similarly, Askew, Kaye, Frasier, Mobasher, Anderson, and Rodriguez (2002) also found that US children who successfully progressed through the Reading Recovery intervention programme maintained their gains in literacy and were performing at a level equivalent to their peers (according to standardized tests) in Fourth Grade, three years later. Such findings were also reflected in Moore and Wade's study (1998), which compared the reading and comprehension levels of 121 students who had received Reading Recovery four to five years earlier with 121 comparison students who were from the same class and who had superior literacy four to five years earlier. It was found that the mean reading-age and comprehension-age of the students who had received Reading Recovery was nearly 12 months and 13 months (respectively) superior to that of the comparison group. A series of other studies also indicate that discontinued Reading Recovery children perform at a level equivalent to their same-aged peers for up to five years post-intervention (e.g., Brown, Denton, Kelly, & Neal, 1999; Askew & Frasier, 1994; Briggs & Young, 2003; Escamilla, Leora, Ruiz, & Rodriguez, 1998; Pinnell, 1989; Rowe, 1995).

However, the long-term follow-up studies presented above (Schmitt & Gregory, 2005; Moore & Wade, 1998; Askew et al., 2002; Brown et al., 1999; Briggs & Young, 2003; Escamilla, Leora, Ruiz, & Rodriguez, 1998; Pinnell, 1989; Rowe, 1995) only included Reading Recovery children who had successfully progressed through the programme and thus, were discontinued. Consequently, these studies do not tell us about the overall effectiveness of Reading Recovery, especially for those children who did not successfully progress through it.

Effects of Reading Recovery three years later

There is evidence of positive programme effects at least up until Second Grade for both discontinued and non-discontinued Reading Recovery children, particularly on those measures tailored to, or designed for, the programme (Hurry & Sylva, 2007; D'Agostino and Murphy, 2004). However, longer-term follow-up studies are less positive., Ruhe and Moore (2005) compared the literacy levels of children from the general population with three groups of Reading Recovery children who received the intervention three years earlier: those who successfully progressed through the programme and were discontinued, those who completed assessments and were recommended for additional assessment, and those with an incomplete series of lessons. It was found that while discontinued children were performing at average levels in reading and writing, children recommended for further assessment and those with an incomplete series of lessons were performing below the average achievement band. This suggests that age-equivalent literacy levels are only achieved by Reading Recovery children who successfully progress through the programme. Hurry and Sylva (2007), following children up till age 10 years, the final year of elementary schooling, found that Reading Recovery gains were only still significant for those in the bottom half of readers at intake. Hurry and Sylva's study evaluated Reading Recovery immediately after its introduction in the UK and therefore with relatively inexperienced Reading Recovery teachers. Evidence on the long-term effectiveness of the programme for 'all' types of Reading Recovery children (regardless of outcome) is relatively sparse, and further research is warranted to investigate this.

There is also a distinct lack of evidence investigating the impact of Reading Recovery on children's later Special Educational Needs status. In one study, O'Connor

and Simic (2002) found that US children who had received Reading Recovery were significantly less likely to have been referred for testing, placed in special education, and/or classified as learning disabled than comparison students who were performing at a literacy level higher than the Reading Recovery children prior to intervention. This suggests that Reading Recovery children are less likely to be identified as requiring some form of special educational provision. However, one of the limitations of this study, acknowledged by the authors, was that these 'Special Educational Needs' statistics were obtained at the end of First Grade at which point definitive assessments for special educational services had not been completed. Consequently, it remains unknown whether the findings would remain over longer periods of time. Indeed, the authors emphasise the need for longitudinal research to investigate whether special education referrals and placements of Reading Recovery children are reduced over longer periods of time.

The durability of gains and the extent to which children avoid the more intensive provision associated with continuing special educational needs not only has a personal impact on the child but also on the cost-effectiveness of intervention, particularly an intensive and large scale intervention such as ECAR and Reading Recovery.

Purpose of the study

In summary, there is a shortage of long-term follow-up data on the durability of the gains made during the Reading Recovery programme. There is also a distinct lack of research investigating the effects of Reading Recovery on children's Special Educational Needs status over longer periods of time. Many of the long-term follow-up studies that exist in the literature can be criticised for only including 'discontinued' Reading Recovery children, for not controlling for any group differences at baseline (D'Agostino

& Murphy, 2004), and for evaluating 'effectiveness' using the Observation Survey as primary evidence rather than some other evaluation tool that is not directly related to the content of the programme (Hiebert, Colt, Catto, & Gury, 1992).

The current evaluation started in 2005, with a sample of six-year-olds in London who had made a slow start in literacy. Children who had received Reading Recovery (n = 87) were compared with similar children who attended London schools where Reading Recovery was not offered (n = 147). The results of this evaluation have been reported as the children reached the end of Year 1 (Burroughs-Lange & Douëtil, 2006) and Year 2 (Every Child a Reader, ECAR, 2008). The children receiving Reading Recovery had made significantly greater progress than the comparison group at both follow-up points. We now report on a further follow-up as the children reach the end of Year 4 (three years post-intervention). We have used children's end of Year 4 National Curriculum Assessments to assess the long-term impact of Reading Recovery on reading, writing, and mathematics, the latter of which would indicate whether effects of Reading Recovery extend beyond literacy.

There were two key questions that were explored in this three-year follow-up study to assess the long-term effectiveness of the Reading Recovery programme:

1. Are there significant group differences between children who received Reading Recovery, children in Reading Recovery schools who did not received Reading Recovery, and comparison children who did not receive Reading Recovery and were not in a Reading Recovery school, in terms of their reading, writing, and mathematics three years later, after controlling for any group differences at baseline?

2. Are there significant group differences between children who received Reading Recovery, children in Reading Recovery schools who did not received Reading Recovery, and comparison children who did not receive Reading Recovery and were not in a Reading Recovery school, in terms of the prevalence of Special Educational Needs?

Method

Participants

At first follow-up in Year 1 (Burroughs-Lange & Douëtil, 2006) there were assessment data on 147 comparison children, 87 children who had received Reading Recovery, and 58 children in Reading Recovery schools who had not received Reading Recovery. At second follow-up in Year 2 (Every Child a Reader, ECAR, 2008), there were Key Stage 1 National Curriculum Assessment data on 140 comparison children, 86 children who had received Reading Recovery, and 51 children in Reading Recovery schools who had not received Reading Recovery. At third follow-up in Year 4 (reported here), there were National Curriculum Assessments data on 241 children: 120 comparison children, 73 children who had received Reading Recovery three years earlier, and 48 children in Reading Recovery schools who had not received Reading Recovery. This represents an attrition rate of 17% and there were similar attrition rates in each of the three groups (18%, 16%, and 17% respectively). Reassuringly, taking the sample as a whole, the children who were untraced did not differ significantly from those traced either in terms of demographic factors or literacy levels. Taking each group separately, on the whole traced and untraced children were similar, but the untraced comparison children scored significantly higher on Book Level (Clay, 2002), $\chi^2(1, N = 146) = 6.01$, p

= .014, and on the Word Recognition and Phonic skills Measure, WRAPS (Moseley, 2003), $t(145) = 2.07$, $p = .04$.

Measures

In Year 1, at baseline, the standard Reading Recovery diagnostic profile was administered, which includes the Observation Survey (Clay, 2002) and the BAS II Word Reading subtest (Elliot, Smith, & McUlloch, 1996). In Year 2, scores were obtained from the BAS II Word Reading subtest, the WRAPS, Progress in English (Kispal, Hagues, & Ruddock, 1994), and Key Stage 1 National Curriculum Assessments for Reading and Writing. In Year 4 (reported here), National Curriculum Assessments for Reading, Writing, and Mathematics were used. Since the comparative analyses reported in this paper only involve data from the tests administered in Year 1 and Year 4, the Year 2 test battery is not described here.

Year 1 Test Battery

The Observation Survey of Early Literacy Achievement

The Observation Survey (Clay, 2002) is the core assessment in the Reading Recovery programme and comprises five key components: in the *Letter Identification task,* children were asked to say out loud all upper and lower case letters including an additional ‘g’ and ‘a’ (which are represented differently) from a list provided. They received one point for every letter that was correctly sounded out and obtained a total score out of 54. In the *Concepts about Print task*, children were assessed for their lower order skills such as their knowledge of messages, directionality, what a letter is, and what a word is etc. They obtained a total score out of 24 on this task. In the *Duncan Word Test* (UK), children were required to read out loud a list of words and obtained a total score out of 23. In the

Writing Vocabulary task, children were asked to write down as many words as they could in ten minutes. The number of correctly written words (including those that were spelled incorrectly, but were deemed accurate in terms of phonics) made up their total score on this task. Lastly, in the *Hearing and Recording Sounds in Words task* (also called dictation), children were asked to write down a short, simple passage that was dictated by the administrator. A total score out of 37 was obtained on this task. In addition to these tasks a Reading Recovery teacher also takes a running record of children's oral reading behaviour as she or he reads a selected text.

The British Ability Scales II Word Reading subtest

Word reading ability was assessed using the BAS II Word reading subtest (Elliot et al., 1996). Children were presented with a list of up to ninety words of increasing difficulty and were required to read out loud as many words as they could from the list provided. Children received one point for every word that was read out accurately and the test was terminated if eight or more errors were made in any one block of ten words.

Table 1 shows the Year 1, baseline characteristics for the comparison children, the Reading Recovery children, and the comparison children in Reading Recovery schools.

Table 1 about here.

It can be seen from Table 1 that the children in this study tended to be economically disadvantaged, with just over half taking free school meals (54%), and having English as an additional language (48%). The majority were effectively non-readers at baseline (50%) not scoring at all on the British Ability Scales II Word Reading subtest and the vast majority (81%) were either not reading or were only able to read the most basic level

books. They did have some skills in place as demonstrated by performance on the Observation Survey of Early Literacy Achievement.

At baseline there were significant demographic differences between the three groups of re-contacted children on uptake of free school meals, $\chi^2(2, N = 241) = 7.21$, $p = .027$, with a higher proportion of the comparison group taking free school meals. There were also significant group differences in baseline literacy, with the comparison children on average scoring lower on the BAS II Word Reading subtest, $\chi^2(2, N = 241) = 10.28$, $p = .006$, and Book Level, $\chi^2(2, N = 238) = 16.84$, $p < 0.001$. The comparison children in Reading Recovery schools were also somewhat weaker than the Reading Recovery group on literacy at baseline, significantly so for the BAS II Word Reading subtest, $\chi^2(1, N = 121) = 4.71$, $p = .03$. However, this measure is rather crude at baseline as half or more children do not score at all. On the more sensitive measures for children at this level, the Observation Survey and the WRAPS, there were no significant differences. Note: for the Observation Survey, scores were standardised to a mean of zero, so positive scores show higher than average scores and negative scores show lower than average scores. Despite the fact that less credence is placed on the BAS II Word Reading subtest at this stage of literacy development, it is regrettable that there were significant group differences at baseline. We acknowledge this as a serious shortcoming of this research so it was essential to control for any group differences at baseline in all subsequent analyses.

Year 4 Test Battery

National Curriculum Assessments for Reading, Writing, and Mathematics

In Year 4, teacher-assessed National Curriculum sublevels for Reading, Writing, and Mathematics were collected for all children. The levels (ranging from below Level 1 to

Level 4a) were converted to National Curriculum point score equivalents for all statistical analyses (see Appendix) following the guidance provided by the UK Department for Education website (http://nationalstrategies.standards.dcsf.gov.uk/node/169521). The National Curriculum Assessments represent an ordinal level of measurement, but such measures are typically analysed using multiple regression, as this form of analysis is sufficiently robust to cope with such data where the sample size is over 200.

To test the validity of these measures we explored the correlations between National Curriculum Assessments at the end of Year 2, the other literacy measurements taken at the same time, and their relationship with National Curriculum Assessments at the end of Year 4 (see Table 2).

Table 2 about here.

It can be seen in Table 2 that in Year 2, the reading measures of the BAS II Word Reading subtest and the Progress in English 7 were highly correlated with National Curriculum Assessment reading (r = .83, n = .112, p < .001 and r = .83, n = .105, p < .001 respectively). These correlations were very similar to the correlations between the reading measures themselves (r = .82, n = .104, p < .001). This supports the validity of the National Curriculum Assessment measures in the current context. Correlations between the BAS II Word Reading subtest in Year 2 and National Curriculum Assessments in reading and writing in Year 4 were also fairly robust (r = .7, n = .95, p < .001 and r = .74, n = .95, p < .001 respectively), as were the correlations between the Progress in English test and National Curriculum Assessments in reading and writing in Year 4 (r = .65, n = .88, p < .001 and r = .66, n = .88, p < .001 respectively). This

supports the Year 4 National Curriculum Assessments as providing valid measures of reading and writing.

Procedure

At the beginning of this longitudinal research, schools from five London boroughs were selected for the Reading Recovery group and schools from five other London boroughs were selected to form the first comparison group. These boroughs were similar in population characteristics and Key Stage 1 achievement levels (Burroughs-Lange & Douetil, 2006). Across the five boroughs, 21 infant and primary schools were identified which in 2005-2006 had adopted ECAR and had a Reading Recovery teacher providing literacy intervention in Year 1. Across the five other boroughs, 21 non-ECAR schools with no Reading Recovery teachers in Year 1 were nominated by the Local Authority as of most concern for high numbers of children with poor performance in literacy. An earlier report (see Burroughs-Lange & Douëtil, 2006) documents that schools were similar in terms of uptake of free school meals, number of children with English as an additional language, school size, and attainment of Year 1 children in September 2005.

In these 42 schools, the standard Reading Recovery diagnostic profile (An Observation Survey of Early Literacy Achievement) and the BAS II Word Reading subtest were administered to assess and identify the eight children considered to be lowest in literacy in their Year 1 classes; these children formed the sample for this evaluation. The selection of children to receive Reading Recovery was based on teachers' recommendations, the students' performances on the Observation Survey and the BAS II Word Reading subtest, but also on their age (older children are often taken first). It was not possible to offer Reading Recovery to all the children in Reading Recovery schools.

Of the 145 children in Reading Recovery schools, 87 received Reading Recovery, 58 did not, and those who did not would make up the second comparison group in this study.

In Year 4, three years after intervention (reported here), teacher-assessed National Curriculum sublevels for Reading, Writing, and Mathematics were collected for all children across the three groups in order to assess the long-term effectiveness of the Reading Recovery programme. Data on the children's special education needs status was also collected from the Pupil Level Census for Spring 2008 (in Year 3) in order to investigate the programme's effectiveness in terms of the prevalence of Special Educational Needs in each of the three groups.

Results

At the end of Year 4, comparison children (n = 120), children who had received Reading Recovery three years earlier (n = 73), and children in Reading Recovery schools who had not received Reading Recovery (n = 48) were compared.

1. *Are there significant group differences between children who received Reading Recovery, children in Reading Recovery schools who did not received Reading Recovery, and comparison children who did not receive Reading Recovery and were not in a Reading Recovery school, in terms of their reading, writing, and mathematics three years later, after controlling for any group differences at baseline?*

Table 3 presents the average National Curriculum Assessment levels for each of the three groups. To provide more precision these are also expressed in National Curriculum point score equivalents, in the form of means and standard deviations. Group differences were tested for statistical significance using multiple regression analyses

controlling for any group differences at baseline (i.e., Observation Survey scores, BAS II Word Reading subtest scores, Book Level scores, free school meals, and English as an additional language).

Table 3 about here.

It can be seen from Table 3 that the Reading Recovery children were still doing significantly better in reading, *Beta* = .21, $t(237) = 3.5$, $p = .001$, and writing, *Beta* = .19, $t(237) = 3.30$, $p = .001$, than the other two groups overall. The differences between the Reading Recovery children and the comparison children in non-Reading Recovery schools were the greatest (reading, *Beta* = .26, $t(189) = 3.9$, $p = .001$, effect size (Cohen's d) = .53 and writing, *Beta* = .23, $t(189) = 3.45$, $p = .001$, effect size (Cohen's d) = .46). The Reading Recovery children had reached an average of 3b in reading and 2a in writing, ahead of the comparison children in non-Reading Recovery schools by just under half a National Curriculum level in reading and a third of a level in writing. The comparison children in Reading Recovery schools were in the middle. They were doing better than the comparison group in non-Reading Recovery schools, though not significantly so (reading, *Beta* = .10, $t(165) = 1.46$, $p = .147$, effect size (Cohen's d) = .11 and writing, *Beta* = .11, $t(165) = 1.50$, $p = .136$, effect size (Cohen's d) = .11). They were not doing as well as children who had received Reading Recovery, but again differences were not statistically significant when assessed using the same multiple regression models (reading, *Beta* = .11, $t(119) = 1.36$, $p = .176$, effect size (Cohen's d) = .22 and writing, *Beta* = .06, $t(119) = .76$, $p = .448$, effect size (Cohen's d) = .12). There were no significant group differences in mathematics.

If only the Reading Recovery children deemed to have successfully graduated from the programme (the 'discontinued' group) were compared with the other comparison groups the differences were even more marked. As would be expected, the mean scores for the Reading Recovery children in Year 4 (N = 62) were higher when the 'referred' children were excluded from the analysis: Reading Mean = 21.91 (SD = 3.78); Writing Mean = 19.65 (SD = 3.35); Maths Mean = 19.4 (SD = 3.32). The differences between the 'discontinued' Reading Recovery children and the comparison children in non-Reading Recovery schools were for reading, *Beta* = .28, $t(188) = 4.24$, $p = .001$, effect size (Cohen's d) = .59, and for writing, *Beta* = .28, $t(188) = 4.28$, $p = .001$, effect size (Cohen's d) = .59.

2. *Are there significant group differences between children who received Reading Recovery, children in Reading Recovery schools who did not received Reading Recovery, and comparison children who did not receive Reading Recovery and were not in a Reading Recovery school, in terms of the prevalence of Special Educational Needs?*

In the UK (although the process is slightly different in Scotland), a child who is making little or no progress at school may receive support through 'School Action' whereby teaching approaches (and the curriculum) are adjusted and targeted particularly in a child's identified area of weakness. The child's teacher and the Special Educational Needs Coordinator (SENCO) will develop an Individual Education Plan (IEP) outlining the child's short-term targets and how these will be met. 'School Action Plus' is similar to School Action, but the SENCO (or equivalent) will seek further advice from external support services (e.g., an educational or school psychologist). If the child fails to benefit

from School Action or School Action Plus they may receive a statement of Special Educational Needs, which is the highest level of support available to children with Special Educational Needs (Special Educational Needs Code of Practice, 2008).

Table 4 presents the Special Educational Needs status for each of the three groups at the end of Year 3.

Table 4 about here.

At baseline (Year 1), very few children in the study had a Statement of Special Educational Need: 5.4% (n = 8) of the comparison group, 11.5% (n = 10) of the Reading Recovery group and 8.6% (n = 5) of the comparison children in Reading Recovery schools. These between-group differences were not statistically significant in Year 1. In the spring of 2008, when the children were at the end of Year 3, data was available through the Pupil Level Census on 267 of the original 293 children. As before, few children had a statement of Special Educational Need, but quite a number were on stages one (School Action) or two (School Action Plus) of the Special Needs Code of Practice (2008), (see Table 4). Overall, Reading Recovery children were receiving significantly less special provision than children in the other two groups, $\chi^2(6, N = 267) = 15.228$, $p = .019$. Fewer Reading Recovery children (as a percentage) were receiving support through School Action Plus or had received a Statement of Special Educational Needs, which are the two highest levels of support available to children with Special Educational Needs. Moreover, a greater percentage of Reading Recovery children were receiving no special provision, although it is noteworthy that the percentage of Reading Recovery children receiving support through School Action (the lowest level of support) was greater than the percentage of comparison children in Reading Recovery schools.

If only the Reading Recovery children deemed to have successfully graduated from the programme (the 'discontinued' group) were compared with the other comparison groups the differences were more marked, $\chi^2(6, N = 246) = 22.194$, $p = .001$. As would be expected, the Reading Recovery children at the end of Year 3 were receiving even less special provision when the 'referred' children were excluded from the analysis: No Special Provision, 65% ($n = 44$); School Action Plus, 6% ($n = 4$); Statement of Special Educational Needs, 0% ($n = 0$), although a greater percentage were on School Action, 29% ($n = 20$).

Discussion

This study set out to examine 1) whether there were significant group differences between children who received Reading Recovery, children in ECAR/Reading Recovery schools who did not received Reading Recovery, and comparison children who did not receive Reading Recovery and were not in an ECAR or Reading Recovery school, in terms of their reading, writing, and mathematics three years later, after controlling for any group differences at baseline, and 2) in terms of the prevalence of Special Educational Needs. These research questions will now be addressed in turn.

Children who received the Reading Recovery intervention three years earlier were still performing at a higher level in reading and writing than comparison children in Reading Recovery schools and comparison children in non-Reading Recovery schools, although only significantly so in the latter. The size of the effect of Reading Recovery versus being in a non-Reading Recovery school was not trivial, (we have used Cohen's d) .53 for reading and .46 for writing. Hattie (2012) suggests that an effect size of .4 is the 'hinge-point' for identifying what is and what is not effective. Half of the influences on

achievement are above this hinge-point. The effect sizes observed in this follow-up study fall above this hinge-point. Almost all of the effects referred to by Hattie are short-term. Thus this can be deemed a very respectable long term effect. It was also found that children who received Reading Recovery were achieving an average National Curriculum level of 3b in reading, which indicates being on track for Level 4 at the end of Key Stage 2 (expected for this age). These findings were anticipated based on the available literature demonstrating that Reading Recovery children maintain their gains in literacy and are able to perform at an age-appropriate level at, and beyond, three years post-intervention (Schmitt & Gregory, 2005; Askew et al., 2002; Brown et al., 1999; Askew & Frasier, 1994; Briggs & Young, 2003; Escamilla et al., 1998; Pinnell, 1989; Rowe, 1995). The findings were also consistent with the literature demonstrating that children who received the Reading Recovery intervention four to five years earlier outperform comparison children on measures of literacy (Moore & Wade, 1998).

One of the short-comings of much of this literature (e.g., Schmitt & Gregory, 2005; Moore & Wade, 1998; Askew et al., 2002; Brown et al., 1999) was that the Reading Recovery evaluation sample consisted of discontinued children only; that is, those who successfully progressed through the programme. In order to evaluate the 'overall' effectiveness of Reading Recovery it is paramount to include all children who received the programme, regardless of outcome. Indeed, evidence from Ruhe and Moore (2005) suggests that gains in literacy are only sustained by those who are successfully discontinued from the programme. Reassuringly, the findings in the present study were based on evaluation data using combined scores from all Reading Recovery children, including both discontinued and non-discontinued children. Therefore, the findings can

be considered to indicate that 'overall' effects of Reading Recovery, regardless of outcome, are still apparent three years post-intervention. However, we acknowledge that the effectiveness of Reading Recovery may reside primarily in those children who are successfully discontinued from the programme (Ruhe & Moore, 2005) who accounted for 85% of the Reading Recovery sample in the present study.

The fact that three years after intervention, the children who had received Reading Recovery were not showing significantly greater progress than the comparison children in Reading Recovery schools does pose some question over the mechanisms at work and consequently cost effectiveness. A similar finding was reported by Hurry and Sylva (2007) for their one year follow-up. The nature of the Reading Recovery model is a systemic one. As part of this, the teacher receives intensive year long training and then returns to his/her school offering the potential for a whole school effect. In the current context Reading Recovery was also embedded in an explicitly whole school approach (ECAR). If the effect is entirely a 'whole school' one then it could be argued that it would be more cost effective to train the teachers, but not to deliver the Reading Recovery intervention. It is also argued that the provision of intensive support for the weakest readers frees up the class teacher to provide more focused attention for other weaker readers. If this is the case then one would expect the greatest effect for those children receiving the intervention and some effect for weaker readers not offered the programme, relative to a control group from non-Reading Recovery schools. In the present study, immediately post intervention the Reading Recovery children were doing significantly better than the comparison children in Reading Recovery schools on the BAS II Word Reading test (Beta = .264, $t(116) = 4.022$, $p = .001$, effect size (Cohen's d)

= .52). The comparison children in Reading Recovery schools were doing significantly better than comparison children in non-Reading Recovery schools, but with a considerably weaker effect (Beta = .2.43, t(149) = 5.000, p = .001, effect size (Cohen's d) = .26). The most plausible interpretation for the results reported here, three years on, is that all of the effects have reduced, and that the gap between the Reading Recovery children and the comparison group in Reading Recovery schools has narrowed as a result. However, if we were to rely on a whole school effect, based on the argument that to implement the one-to-one intervention was not cost effective, in the present study there would be no significant effects to report. Only the comparison between Reading Recovery children and the comparison group in non-Reading Recovery schools reached statistical significance. Perhaps the sustained gains observed here are acombination of one-to-one intervention embedded within a consistent whole school programme. Research on reading interventions converges in concluding that they should be consistent with classroom instruction (Foorman & Moats, 2004).

Moreover, children who received Reading Recovery were significantly less likely to be receiving some form of Special Educational Needs provision than both comparison groups. These findings were consistent with O'Connor and Simic (2002) who found that Reading Recovery children were significantly less likely to have been referred for testing, placed in special education, and/or classified as learning disabled than comparison children at the end of First Grade. The findings from this study suggest that this pattern of results remains over longer periods of time.

While this study offers some unique insights regarding the Reading Recovery programme's long-term effectiveness in terms of sustainability of gains in literacy and

prevalence of special education referrals and placements, it does have some limitations. For instance, random allocation is considered to be one of the hallmarks of rigorous intervention research (Chapman, Greaney, & Tunmer, 2007); yet, few studies have used some form of random allocation (Shanahan & Barr, 1995). In this study, the selection of children to receive Reading Recovery was based on teachers' recommendations, the students' performances on the Observation Survey and the BAS II Word Reading subtest, and on their age, with older children being taken first. As a likely result of this, significant group differences were found at baseline on uptake of free school meals, the BAS II Word Reading subtest, and Book Level. Although these differences were controlled for in all subsequent analyses, it is regrettable that better-matched comparison groups were not established at the outset of this study. However, it should be noted that historically it has proven very difficult to evaluate the effectiveness of the Reading Recovery programme due to policies regarding student selection and barriers to locating an equivalent comparison group (D'Agostino & Murphy, 2004). It is also noteworthy that in a review of the literature on literacy and numeracy interventions, it was found that only 34 of 806 relevant articles (4.22%) used random allocation (Seethaler & Fuchs, 2005).

In summary, this study does not address many other popular challenges in the literature concerning Reading Recovery, such as the programme's efficacy in comparison to other interventions, its cost effectiveness, and whether it supports those who have phonological difficulties. However, it does provide some much-needed evidence on the durability of the gains made resulting from the Reading Recovery programme and the effects of Reading Recovery on children's Special Educational Needs status over longer periods of time. Based on the findings reported here, it can be concluded that children

who receive the Reading Recovery intervention perform at an expected level for their age three years after intervention, at a level higher than those not attending Reading Recovery schools. The findings also suggest that children who receive the Reading Recovery intervention are significantly less likely to be on some level of the Special Educational Needs Code of Practice (2008) at the end of Year 3 than children who had not received Reading Recovery, in non-Reading Recovery schools. This has both emotional implications for the child and their family, but also financial implications for the school and the Local Authority. Finally, it is important to acknowledge that these effects were observed in a few standard Reading Recovery schools, providing a programme that was being taught in many schools all over the country. It is one thing for researchers to demonstrate that a particular approach is effect in remediating reading difficulties in a research context, where programme fidelity is highly controlled. It is quite another to observe significant effects in a working, nationally implemented model.

References

Askew, B. J., & Frasier, D. F. (1994). Sustained effects of Reading Recovery intervention on the cognitive behaviors of second grade children and the perceptions of their teachers. *Literacy, Teaching, and Learning: An International Journal of Early Literacy, 1*(1), 87–107.

Allington, R. L. (1984). Content coverage and contextual reading in reading groups. *Journal of Reading Behavior, 16*, 85-96.

Askew, B. J., Kaye, E., Frasier, D. F., Mobasher, M., Anderson, N., & Rodriguez, Y. G. (2002). Making a case for prevention. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing, 6*(2), 43–73.

Biemiller, A. (1977-78). Relationships between oral reading rates for letters, words, and simple text in the development of reading achievement. *Reading Research Quarterly, 13,* 223-253.

Briggs, C., & Young, B. K. (2003). Does Reading Recovery work in Kansas? A retrospective longitudinal study of sustained effects. *The Journal of Reading Recovery, 3*(1), 59–64.

Brown, W., Denton, E., Kelly, P., & Neal, J. (1999). Reading Recovery effectiveness: A five-year success story in San Luis Coastal Unified School District. *ERS Spectrum: Journal of School Research and Information, 17*(1), 3–12.

Burroughs-Lange, S. & Douëtil, J. (2006). Evaluation of Reading Recovery in London Schools: Every Child a Reader 2005-06. http://www.everychildareader.org/pub/index.cfm

Chall, J. S. (1983). *Stages of Reading Development*. New York: McGraw-Hill.

Chapman, J., Greaney, K., & Tunmer, W. (2007). How well is Reading Recovery really working in New Zealand? *New Zealand Journal of Education Studies, 42*(1), 17-29.

Chapman, J. W., Tunmer, W. E., & Prochnow, J. E. (2001). Does success in the Reading Recovery program depend on developing proficiency in phonological processing skills? A longitudinal study in a whole language instructional context. *Scientific Studies of Reading, 5*(2), 141-176.

Clay, M. M. (1967). The reading behaviour of five year old children: A research report. *New Zealand Journal of Educational Studies, 2,* 237-248.

Clay, M. M. (1985). *The early detection of reading difficulties: A diagnostic survey with*

recovery procedures (3rd ed.). Auckland: Heinemann.

Clay, M. M. (2002). *An Observation Survey of early literacy achievement* (2nd ed.). Auckland: Heinemann.

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

Elliott, P., Smith, P. & McCulloch. K, (1996) *British Abilities Scales*. Windsor: NFER-Nelson.

Escamilla, K., Loera, M., Ruiz, O., & Rodriguez, Y. (1998). An examination of sustaining effects in Descubriendo la Lectura programs. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing, 3*(2), 59–81.

Every Child a Reader (2008). Every Child a Reader: The results of the third year… http://readingrecovery.ioe.ac.uk/reports/documents/Every_Child_a_Reader_-_results_of_the_third_year.pdf

Foorman, B. R., Moats, L.C. (2004). Conditions for sustaining research-based practices in early reading instruction. *Remedial and Special Education, 25(1),* 51- 60.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning.* London: Routledge.

Hiebert, E. H., Colt, J. M., Catto, S. L., & Gury, E. C. (1992). Reading and writing of first-grade students in a restructured Chapter 1 program. *American Educational Research Journal, 29*(3), 545–572.

Hurry, J. (1996) What's so Special about Reading Recovery *The Curriculum Journal 7 (1),* 93-108.

Hurry, J. & Sylva, K. (2007) Long-term outcomes of early reading intervention. *Journal*

of Reading Research, 30, 227-248.

Juel, C. (1988). Learning to read and write: a longitudinal study of fifty-four children from first through fourth grade. *Journal of Educational Psychology, 80,* 437-447.

Kispal, A., Hagues, N. & Ruddock, G. (1994) *Progress in English7.* Windsor: NFER-Nelson.

Moore, M., & Wade, B. (1998). Reading and comprehension: A longitudinal study of ex Reading Recovery students. *Educational Studies, 24*(2), 195-203.

Moseley, D. (2003). *Word Recognition and Phonic Skills* (2nd ed.). Abingdon: Hodder & Stoughton.

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development/National Institutes of Health.

O'Connor, E. A., & Simic, O. (2002). The effect of Reading Recovery on special education referrals and placements. *Psychology in the Schools, 39*(6), 635-646.

Pinnell, G. S. (1989). Reading Recovery: Helping at risk children learn to read. *Elementary School Journal, 90*(2), 161–184.

Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly, 29,* 9–39.

Rowe, K. J. (1995). Factors affecting students' progress in reading: Key findings from a longitudinal study. *Literacy, Teaching and Learning: An International Journal of Early Literacy, 1,* 57–110.

Ruhe, V., & Moore, P. (2005). The impact of Reading Recovery on later achievement in reading and writing. *ERS Spectrum, 23*(1), 20–30.

Schmitt, M. C., & Gregory, A. E. (2005). The impact of an early literacy intervention: Where are the children now? *Literacy Teaching and learning: An International Journal of Early Reading and Writing*, *10*(1), 1-20.

Schwartz, R. M., Hobsbaum, A., Briggs, C. & Scull, J. (2009). Reading Recovery and evidence-based practice: A response to Reynolds and Wheldall (2007). *International Journal of Disability, Development and Education, 56*(1), 5-15.

Seethaler, P., & Fuchs, L. (2005). A drop in the bucket: Randomized contact trials testing reading and math interventions. *Learning Disabilities Research and Practice, 20,* 98-102.

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

Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360-406.

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

Wasik, B. A. & Slavin, R. E. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly, 28*(2), 179-200.

Appendix.

National Strategy Sublevels: point score equivalents using the a, b, c indicator

Level	Point score
1c	7
Level 1	9
1b	9
1a	11
2c	13
Level 2	15
2b	15
2a	17
3c	19
Level 3	21
3b	21
3a	23
4c	25
Level 4	27
4b	27
4a	29

Note. a = strong level, b = sound level, c = weak level.
Caution should be used with sublevels as the National Curriculum level was designed to indicate representative attainment at the end of a Key Stage; a sublevel only gives an indication of the certainty of this achievement, but can be useful in identifying progress.

Table 1.

Baseline characteristics of the comparison children not in Reading Recovery schools, the Reading Recovery children, and the comparison children in Reading Recovery schools re-contacted in Year 4

	% FSM	% EAL	% No BAS score	% ≤ Book Level 1	OS z-score Mean (SD)	WRAPS Mean (SD)
Comp. not in RR schools (*n* = 120)	63%	55%	57%	91%	-.030 (.95)	10.6 (5.9)
RR children (*n* = 73)	44%	47%	34%	72%	.137 (.88)	11.6 (6.3)
Comp. in RR schools (*n* = 48)	52%	35%	54%	69%	-.032 (1.25)	12 (10.2)
Total (*n* - 241)	54%*	48%	50%**	81%***	.020 (1)	11.2 (7.1)

Note. RR, Reading Recovery; FSM, Free School Meals; EAL, English as an Additional Language; BAS, British Ability Scales II Word Reading subtest; OS, Observation Survey; WRAPS, Word Recognition and Phonics skills.
*p<.05, **p<.01, ***p<.001

Table 2.

Spearman correlations between National Curriculum Assessments and other literacy measures: Non-RR comparison children only

			Year 2			*Year 4*		
	BAS	WRAPS	PiE7	NC: Reading	NC: Writing	NC: Reading	NC: Writing	NC: Maths
Year 2								
BAS								
WRAPS	.887***							
PiE7	.821***	.847***						
NC: Reading	.833***	.808***	.827***					
NC: Writing	.733***	.708***	.763***	.838***				
Year 4								
NC: Reading	.704***	.607***	.648***	.740***	.602***			
NC: Writing	.739***	.702***	.659***	.681***	.637***	.763***		
NC: Maths	.482***	.447***	.427***	.554***	.522***	.589***	.589***	

Note. RR, Reading Recovery; BAS, British Ability Scales II Word Reading subtest; PiE7, Progress in English; NC, National Curriculum.
*p<.05, **p<.01, ***p<.001

Table 3.

Mean National Curriculum levels and point score equivalents of the comparison children not in Reading Recovery schools, the Reading Recovery children, and the comparison children in Reading Recovery schools re-contacted in Year 4

		Reading	Writing	Maths
Comp. not in RR schools (*n* = 120)	NC level	2a	2b	2a
	Mean point score	18.21	16.43	18.39
	SD	5.20	4.55	4.49
RR children (*n* = 73)	NC level	3b	2a	2a
	Mean point score	20.14	18.75	18.92
	SD	4.27	3.94	3.44
Comp. in RR schools (*n* = 48)	NC level	3c	2a	2a
	Mean point score	19.21	17.38	17.46
	SD	6.25	5.39	5.19

Note. RR, Reading Recovery; NC, National Curriculum.

Table 4.

Special Educational Needs status of the comparison children not in Reading Recovery schools, the Reading Recovery children, and the comparison children in Reading Recovery schools at the end of Year 3

	No Special Provision	School Action	School Action Plus	Statement of SEN
Comp. not in RR schools (*n* = 134)	48% *n* = 64	32% *n* = 43	19% *n* = 25	1.5% *n* = 1
RR children (*n* = 81)	61% *n* = 49	26% *n* = 21	12% *n* = 10	1.2% *n* = 1
Comp. in RR schools (*n* = 52)	42% *n* = 22	21% *n* = 11	29% *n* = 15	7.7% *n* = 4

Note. RR, Reading Recovery; SEN, Special Educational Needs.