A Reading Recovery Comparison Study: Supporting a New Implementation in Scotland

Gillian Gourlay, East Renfrewshire Council.

Sinead Harmey, UCL Institute of Education.

Abstract

Many children in Scotland fail to reach age-expected levels in literacy by the time they leave primary education. Reading Recovery (RR) is a well-researched early literacy intervention that has been recently introduced to address the poverty-related attainment gap in Scottish schools. The purpose of this paper is to (a) describe a study which explored the effectiveness and impact of RR on the literacy achievement of children experiencing literacy difficulties, compared to peers who did not receive the intervention and (b) describe how the results have been used to support the implementation of RR in a new implementation in two districts in Scotland. Results from the study demonstrated that all children in the study made gains, though the RR intervention yielded more positive effects. The children who received the RR reached age expected levels by the end of the intervention and they continued to make gains greater than those of the children who did not receive RR. How these results were used and implications for practice are discussed.

Keywords: Early intervention, literacy, Reading Recovery

A Reading Recovery Comparison Study: Supporting a New Implementation in Scotland

Reading Recovery® (RR) is a short-term, early literacy intervention for the lowest achieving children in literacy (see Clay, 2016 for a full description). The intervention involves the close observation of a child, in a one-to-one tutorial situation, thinking about what that child needs to be able to do next and what they need to change in the way they are reading and writing in order to achieve success (Douëtil, 2005). It is expected that when the RR intervention is put in place, children will reach age expected levels after about 12-20 weeks of instruction (Clay 1985; cited in Clay, 2015). Burroughs-Lange and Douëtil (2007) suggested that schools could have the power to enable almost every child to read and write at age-expected levels if they were to provide them with access to RR after one year of formal schooling.

RR has been implemented across the world. In Europe, RR is implemented in England, Ireland, the Channel Islands, Denmark, Malta, and Scotland. RR had been implemented in Scotland around about 1997 but, due to unforeseen circumstances, had to stop around about 2001 (Douëtil, Hobsbaum & Maidment, 2013). Therefore, RR is relatively new to the context. Within recent years in Scotland there has been an emphasis on providing equity in education as part of '*The Scottish Attainment Challenge'* (Scottish Government, 2018a). Schools are provided with Pupil Equity Funding (PEF) (Scottish Education Council, 2018) for children who are entitled to free school meals but they must be able to show the impact that their spending has had on attainment. Three authorities decided to invest in Reading Recovery as an evidence-based intervention to support young children's literacy learning. Consequently, in 2017-2018, three educators from Scotland undertook the yearlong Masters level programme of study to become accredited as RR Teacher Leaders. RR is now in its third year in three authorities in Scotland and its first year in 3 further authorities. Given the importance of demonstrating the impact of

government spending on pupils' literacy achievement, the purpose of this article is to (a) describe the results of a year-long comparison study conducted during the second year of the implementation of RR in two districts in Scotland and (b) demonstrate how the results have been used to promote the implementation of RR in the district. Implications for practice are discussed in terms of how the results of the study were communicated at local level and what further research could be carried out to confirm the long-term effects of the intervention

Background

There is no shortage of research on RR; it is one of the most widely researched interventions (D'Agostino & Harmey, 2016) and, according to Cunningham and Allington (1994; cited in Lyons, 2003), no other programme "has ever come close to achieving the results demonstrated by Reading Recovery" (p. 2). Bodman (2019) explained that RR data is collected and analysed annually and from this, the short-term impact is evident, with just over eight out of ten children reaching age-expected levels by the end of the intervention.

D'Agostino and Harmey (2016) carried out an international meta-analytic review of comparison groups in the United States and other nations that were implementing the RR intervention. They found that RR had a positive effect on literacy achievement. A What Works Clearinghouse [WWC] (2013) intervention report also claimed that RR has 'positive effects on general reading achievement and alphabetics as well as potentially positive effects on reading fluency and comprehension for beginning readers' (pg. 1). If the intervention report were updated to include findings from the recent final report on the i3 Scale Up of RR (May, Sirinides, Gray & Goldsworthy, 2016), the WWC ratings would include positive findings in all four beginning reading outcome domains (Schwartz, 2018).

When addressing the issue of sustainability, D'Agostino & Harmey (2016) identified short-term gains for RR children but argued that there was a lack of follow-up studies to confidently report on the long-term effects of the intervention, suggesting that long-term effects need to be further investigated. Chapman and Tunmer (2019) suggested that "RR is of some benefit for some students, at least in the short term" (p. 259). They also questioned whether all children who receive the intervention make gains and whether any long-term gains are evident (Chapman & Tunmer, 2019). A recent comparison study in England, however, found that children who received RR made significant gains in all areas compared to a comparison group who did not have RR in their school (Burroughs-Lange and Douëtil, 2007). A ten-year follow-up study on this research found that the initial substantial effects of RR were sustained long-term (Hurry & Fridkin, 2018). The RR group had "significantly higher overall GCSE (General Certificate of Secondary Education) point scores than the comparison group" (Hurry & Fridkin, p. 2).

Given the relative newness of RR to Scotland, none of the research cited in this section was conducted in the Scottish context. While empirical evidence at international and national levels is crucial, if innovations are to embed at local level it is vitally important to have local data. Schildkamp (2018) argued that data use can lead to school improvement but that the data must be contextualized and speak to the stakeholders who collected the data.

Measuring Literacy Achievement in RR

As readers of *The Journal of Reading Recovery* will know, within RR there are a series of initial observation tasks, taken from Clay's (2013) *An Observation Survey of Early Literacy Achievement* (OSELA). These tasks are designed to identify the lowest performing children and provide valuable information about their strengths and difficulties in literacy learning (Clay, 2016). This check takes place after a child's first year in formal education, around the age of six (Clay,

2016). After administering the tasks with children in the age band, the lowest performing children are selected to be a part of the intervention (Clay, 2016).

The tasks within the OSELA are designed to be similar to tasks that children are involved in within the classroom and include letter identification, hearing and recording sounds in words, writing vocabulary, word reading, concepts about print and an assessment of reading continuous text using a Running Record (Clay, 2013). Clay (2013) suggested that the tasks used within the OSELA aim to provide RR teachers with information that will improve instruction by involving them in the detailed observation of individual children, which then allows them the opportunity to create a series of lessons that start with what the child already knows. It is argued, however, that the tasks are "closely aligned to skills taught in Reading Recovery and are considered inherent to the treatment" (Slavin *et al.*, 2011, p.6) – in other words, the suggestion might be that they do not yield objective achievement data. Certainly, the OSELA test measures have larger effect sizes than treatment-independent test measures; but, D'Agostino and Harmey (2016) suggested that this may indicate that the OSELA is more sensitive to change. This may explain the results of Slavin et al., (2011) on their research on the achievement outcomes of alternative approaches for struggling readers, as they did not include studies which used OSELA measures for assessment. They found that although the outcomes for RR were positive, they were not as positive as may have been expected (Slavin et al., 2011). Thus, within the United Kingdom, the British Ability Scale Word Reading Test 3 (BAS 3) (Elliott and Smith, 2011) is used as an external measure to assess a child's word reading age.

Study Purpose and Research Questions

It is important to continue to collect data on the efficacy of RR, particularly at local level and to use a variety of measures to measure efficacy. At a local level, 'A Curriculum for

Excellence' (ACfE) was introduced into the Scottish Education system in 2010. This was "designed to achieve a transformation in education in Scotland by providing a coherent, more flexible and enriched curriculum" (Education Scotland, 2019). Through ACfE there is a national approach taken to improve the wellbeing of children and young people. This is known as *Getting it Right for Every Child* (GIRFEC) (Education Scotland, 2017). The GIRFEC approach believes in the use of early intervention to support children and "puts the best interests of the child at the heart of decision making," (Education Scotland, 2017). Recent figures show that many children in Scotland are not achieving age-expected levels in literacy and these figures appear to increase as children get older. For example, in 2017/2018, 19% of children were not achieving age-expected levels in reading by the end of Primary (P) 1, their first year in formal education. In the same year, there were 21% of children not achieving age-expected levels in reading at the end of P7, their final year in primary education (Scottish Government, 2018b). These figures confirm the suggestion that most students continue to make progress but many of those who are falling behind continue to do so over time and the gap continues to widen (Clay, 2016).

As described in the introduction to this article, within recent years in Scottish Education there has been a big emphasis on providing equity in education as part of "The Scottish Attainment Challenge" (Scottish Government, 2018a). Schools are provided with Pupil Equity Funding (PEF) (Scottish Education Council, 2018) for children designated as in poverty but must demonstrate impact on achievement. Reading Recovery (RR) was viewed as potentially helpful in providing the solution to ensuring that PEF is spent with a view to achieving equity in education. The RR intervention allows for attainment to be tracked and impact to be seen; however, there are no comparison studies within the Scottish context which look at this early literacy intervention. It is imperative, therefore, to conduct this type of research within the Scottish context. It should be

noted that this was a small-scale exploratory study which would provide schools and local authorities with preliminary information that could help to inform them on the effectiveness of RR as an early literacy intervention that could be implemented as a way of addressing the poverty related attainment gap in Scottish schools.

The research questions were:

- Is there a significant improvement in literacy scores for the group who received RR in Autumn of P2 and does this continue to improve over time?
- Is there a significant difference in literacy scores between the RR group, who received RR in Autumn of P2, and a low comparison group, who did not receive RR in the Autumn of P2, in initial assessments and after the RR intervention?

Method

Participants

The selection of participants was "logically influenced by the research questions and the research design" (Ogier, 2002, p. 49). To ensure the study was reliable and valid, a large sample group was offered the opportunity to participate in the study. This sample was taken from the population, which included all schools with teachers that were part of initial professional development groups for one teacher leader. After ethical clearance was obtained for this study, each of these schools was invited to be a part of the research; provided consent was gained from head teachers, RR teachers, parents and the children themselves. During recruitment, schools and teachers were approached. Written informed consent was obtained from head teachers, teachers and parents. After consent was received from parents, the children were also asked to consent to being a part of the study.

From a possible 17 schools, 8 volunteered to be a part of the study. After parental and child consent was received, there were a total of 46 children included in the study. Table 1 shows the breakdown of participants from each school, showing how many were receiving RR during the study and how many were tested not instructed (TNI) (screened for RR but did not receive the intervention).

[Table 1 about here]

Context

Each of the schools involved in the study were invited to take part due to the fact that they currently had a member of staff participating in IPD training. IPD is a year-long training programme which requires teachers to attend fortnightly (every other week) sessions with the outcome being that they become an accredited RR teacher. Seven of the eight schools involved within the study had one RR teacher being trained within an IPD group and one school had two teachers participating in IPD training.

Schools in Scotland are ranked 1–10 on the Scottish Index of Multiple Deprivation (SIMD), with 1 being schools in the most deprived areas and 10 being schools in the most affluent areas. The eight schools involved are all described as being multi-denominational schools. These schools span two local authorities and four of them, schools B, H, K and N, are located within deciles 1 – 3 (see Table 1) on the SIMD, which ranks them as being within a highly deprived area (Scottish Government, 2017). From these schools, 18 children out of the 22 involved in the study were eligible for Pupil Equity Funding (PEF). One school, school G, is located in decile 5 on the SIMD, suggesting this is an area of moderate deprivation. School G had no participants that were eligible for PEF. The final three schools, C, I and L, are located within deciles 9 and 10, suggesting they are in more affluent areas. Although school I is located within an area classed as affluent, the

catchment of this school takes in children within decile 3, classed as highly deprived, and decile 4, described as being moderately deprived. This school had 4 participants that were eligible for PEF.

Measures

There were four tests used to assess the progress of the children during this study. Two of these came from the OSELA (Clay, 2013): running records and writing vocabulary. The third test is the British Abilities Scale 3 (BAS3) Word Reading Test (Elliott & Smith, 2011) which is commonly used as an assessment tool in RR. A Single Word Reading Test (SWRT) (GL Education, 2019) was used at the final test time as an external measure. This test was only carried out at time 3 due to issues surrounding personnel and time constraints. Using this test as an external measure at time point 3 would provide an appropriate comparison without placing any undue stress on the participants.

A running record is an assessment of oral reading of continuous text. Teachers are trained in carrying out running records to ensure that there is a standard way of recording the behaviours that occur as the child reads (Clay, 2013). This ensures the reliability of this measure as reading levels are "obtained according to common practice" (Clay, 2013, p.59). The writing vocabulary test is a timed text production task, measuring a child's ability to produce accurately spelled words. This test is deemed as more reliable than a spelling test which is constrained to a particular set of words and cannot be "generalised to the writing of words outside the set of words" (Clay, 2013, p.115). The BAS3 Word Reading Test assesses a child's word reading ability. Rasch scaling is adopted, allowing the examiner to minimise the number of items to be administered, depending on the child's responses, which means the test is completed quickly (Swinson, 2013), not putting any undue stress on the child. The results of this test are standardised on a sample of almost 1500

British children, including children from a variety of ethnic backgrounds (Swinson, 2013). The SWRT provides a measure of a child's word reading skills and contributes to an assessment of reading achievement by providing a reading age (Foster, 2007). This test is designed to monitor a child's word reading skills and is fully standardized providing standard age scores and reading ages (Foster, 2007).

Data Collection and Analysis

The test data were collected at three points in time throughout the study. These time points were: August 2018, January 2019 and March 2019. All RR and TNI children were assessed at these time points within a two-week window to ensure results provided a reliable comparative estimate. These tests were carried out with both RR and TNI children, and were compared using an independent samples t-test. This would test the significance of the difference between the scores (Ogier, 2002) of the RR and TNI children. At time point 3, The TNI group was split further as 12 of these 22 children became part of RR cohort 2 (RR2), beginning RR in the Spring of P2. Essentially, the lowest performing children were removed from the TNI group (leaving 10 children who at two points in time were identified as experiencing difficulties but were not the lowest in the age cohort. This group is labelled the TNI- group). Therefore, the most meaningful results were the comparison of the TNI group and the RR group at time 2.

Findings

Change over time in literacy achievement (RR Group)

To investigate the first research question, all test results were entered into SPSS statistical software (version 25) and an independent samples t-test was conducted to examine the data. Table 2 shows the means and standard deviations on each of the tests carried out by the RR group at three points in time. Means are presented and converted to reading ages which are reported in years

and months. The mean book level on entry to the RR intervention was 1.83 (*SD* = 2.46), which equates to a reading age of between 4:0 and 5:0 years. This increased to a mean level of 15.67 (*SD* = 3.42) at point 2, which equates to a reading age of between approximately 6:0 and 6:6. At time 3, this continued to increase further to a mean level of 17 (*SD* = 4.66), equating to a reading age of between 6:6 and 7:0.

[Table 2 about here]

A similar pattern can be seen with the mean number of words written during the writing vocabulary test. These scores were compared against stanine tables (Clay, 2013), where an average stanine band for these ages would be 4 or 5. A score of 6.83 at time 1 equates to a stanine of 1 for both age bands, which is significantly below average (Clay, 2013). A score of 35 at time 3 would equate to a stanine of 6 for the lower age band and 5 for the higher age band, suggesting that the children had made progress which helped them to reach average levels or above. Using the BAS3 conversion tables (Elliot and Smith, 2011), the mean score at time 1 (*M* =5.21, *SD* = 4.03) equates to a word reading age of 5:4. By time 3 this had increased to a word reading age of 6:4 (*M* = 30.88, *SD* = 10.97), showing a gain of 1 year.

Comparison of literacy scores between RR and TNI groups

To ascertain if there was a significant difference in literacy scores between both groups in initial assessments, an independent samples t-test was used to examine data gathered from initial assessments. Table 3 shows the means and standard deviations of both the RR and TNI groups on the initial assessments at time 1. The mean for the initial assessment of book level was 1.83 (*SD* = 2.46) for the group receiving RR and 2.86 (*SD* = 3.59) for the TNI group. The mean scores for the writing vocabulary initial assessment for the RR group was 6.83 (*SD* = 4.31) and 8.82 (*SD* = 7.08)

for the TNI group. On the BAS3 word reading test the mean was 5.21 (*SD* = 4.03) for the RR group and 7.91 (*SD* = 7.35) for the TNI group.

[Table 3 about here]

A *'p'* value of less the 0.05 would mean that a difference in means is statistically significant (Robson and McCartan, 2016). The results revealed that there was no statistically significant difference in the mean scores of the two groups in initial tests on book levels ($t_{(44)}$ = -1.14, p = 0.26), writing vocabulary ($t_{(44)}$ = -1.16, p = 0.25) or the BAS3 word reading test ($t_{(44)}$ = -1.56, p = 0.13). When comparing the standard deviations (*SD*) of both groups it can be seen that TNI group have larger *SD*'s on all test results compared to the RR group. This means there is more of a variation in scores for the TNI group.

In order to determine whether the means of the RR and TNI group change over time, an independent samples t-test was conducted to compare assessment results at two further points in time. Table 4 shows the means of each of the three tests at two further points in time. The effectiveness of the RR intervention is best reflected by the comparison of the RR and TNI group at time 2, since by time 3, 12 of the 22 children in the TNI group had begun to receive RR lessons.

[Table 4 about here]

Book Level. At time 2, book level had increased to a mean of 15.67 (*SD* = 3.42) for the RR group and 8.55 (*SD* = 7.20) for the TNI group, equating to a reading age of 6:6 and 5:6 respectively. The results revealed that the difference between the levels of the two groups was statistically significant; ($t_{(44)}$ = 4.34, p < 0.01). At time 3, these levels had further increased to a mean book level of 17 (*SD* = 4.63) for the RR group and 12.68 (*SD* = 6.58) for the TNI group, equating to approximate reading ages of 6:6 and 5:9 respectively. Again, the results revealed that this difference in levels was statistically significant; ($t_{(44)}$ = 2.59; p < 0.01). The gains for the RR

group between time 2, the end of the intervention period, and time 3 shows continued growth in their classroom context by children who completed the RR intervention. A comparison of change over time in book level, for both the RR and TNI groups, can be seen in Figure 1.

[Figure 1 about here]

Writing Vocabulary. At the second time point the RR group increased their number of words written in 10 minutes to a mean score of 32.83 (*SD* = 11.54). The mean score for the TNI group was 21 (*SD* = 11.17). This difference in results was again, statistically significant; ($t_{(44)}$ = 3.53; $p < 0.01$). By final assessment (time 3) the mean number of words written was 35 (*SD* = 8.88) for the RR group and 27.14 (*SD* = 14.05) for the TNI group. The mean scores at time 3 equate to a stanine of 5 or 6 for the RR group and 4 or 5 for the TNI group. This difference in results was found to be statistically significant; ($t_{(44)}$ = 2.29; p = 0.03). Change over time in results from the Writing Vocabulary test from time 1 to time 3 can be seen in Figure 2.

[Figure 2 about here]

British Ability Scale Word Reading Test 3 (BAS3). At time 2 the mean scores of the BAS3 test had increased to 27.25 (*SD* = 10.94) for the RR group and 18 (*SD* = 13.63) for the TNI group. The results revealed that there was a statistically significant difference between the scores of the two groups; ($t_{(44)}$ = 2.55; $p < 0.01$). By time 3 there was a further increase in mean to 30.88 (*SD* = 10.97) for the RR group. The TNI group mean at this time was 25 (*SD* = 14.38). The mean results at time 3 equate to a word reading age of 6:4 for the RR group and 6:1 for the TNI group. This difference in results was not found to be statistically significant; ($t_{(44)}$ = 1.57; p = 0.13). The BAS3 test results of both groups are compared in Figure 3, where change over time can be observed.

[Figure 3 about here]

Word Reading. At the final time point, the GL Assessments Single Word Reading Test (SWRT) (Foster, 2007) was conducted as an external measure with both groups. Table 4 shows that the mean score in the SWRT test was 17.54 (*SD* = 6.56) for the RR group and 13.55 (*SD* = 8.18) for the TNI group. Using the SWRT conversion tables (Foster, 2007) this equates to a reading age of 6:6 for the RR group and between 6:0 and 6:3 for the TNI group (Foster, 2007). The difference in the results for this test, however, was not found to be statistically significant ($t_{(44)}$ = 1.84, p = 0.07).

The time 2 assessment captures the progress of the RR group across the intervention period. At this point the lowest performing students in the TNI group entered the intervention. As can be seen in Table 5, the children remaining in the TNI- group score about the same as the initial RR group at time three, indicating that the some TNI students have been able to make progress in their classroom context. The slightly larger standard deviations for the TNI- group indicate larger variation in scores for this group than the RR or RR2 groups. This suggests that some low performing students remain in the TNI- group. Schools might further examine this variation to ensure that sufficient RR services are available to support all students who are not able to make progress with the classroom based literacy instruction.

(Table 5 about here)

Discussion

Improvement in literacy scores of the RR group to age-expected levels

It was found that the mean levels on all 3 test measures of literacy had increased at each time point. This is evidence that the RR group made gains in literacy which continued over time. As children who are part of the RR intervention are around the age of six years old (Clay, 2016), it can be concluded that this group of children have reached age-expected levels, as a book level

of approximately 16 at time two equates to a reading age of between 6:0 and 6:6. The group then continued to make progress after the intervention had finished to a reading age of between 6:6 and 7:0. This is in line with many studies which claim that accessing RR can enable almost every child to read and write at age-expected levels (Clay, 1991; Cunningham & Allington, 1994; cited in Lyons, 2003; Schmitt *et al.,* 2005; Burroughs-Lange & Douetil, 2007; D'Agostino & Harmey, 2016; Hurry & Fridkin, 2018; Bodman, 2019). It demonstrates that the investment in RR was having a direct positive impact on the children selected as required by the PEF funding guidelines. As suggested in the introduction, this information can help to inform head teachers on the effectiveness of the intervention and provides empirical context specific evidence that implementing the intervention can help to address the poverty related attainment gap in Scottish schools.

Higher gains in literacy for the RR group

An independent samples t-test was used to compare results of the RR group, who received RR in Autumn of P2, and the TNI group who did not. When the results were analyzed it was noticed that, as expected, the TNI group had higher mean scores at time one. This would explain why this group were not selected to receive RR at this time, as Clay (2016) explains that RR targets the lowest performing children. This counters Chapman and Tunmer's (2019) claim that the pupils who are accepted into the programme are not actually the lowest performing and that the selection process is manipulated to make results look better. By time two the group who received RR made statistically significant gains in all areas compared to the TNI group and outperformed them on all three tests. It has been suggested that RR provides short term gains for those that are part of the intervention (Burroughs-Lange & Douetil, 2007; Reynolds, Wheldall & Madeline, 2009; What

Works Clearinghouse, 2013; D'Agostino & Harmey, 2016; Bodman 2019) and this is evident in these results.

We argue that this study makes a strong case for the need for RR in this area. Had there been enough places, many of TNI group would certainly have qualified for RR given that the average book level for this group was well below average and that their average reading age was 5 years and 7 months. This was confirmed by the fact that 12 of the 22 children went on to be selected for RR later in the school year (RR2 group in Table 5). This implies that they did not profit from other or mainstream instruction in the interim. These findings provide useful feedback to schools and authorities about what full implementation of RR might look like and , in this case, for fulfilling the aims of the *Getting it Right for Every Child Policy* (Education Scotland, 2017) which aims to provide early intervention to those in need. Tracking children who are not initially selected for RR on the first round is important for comparative purposes and provides useful data on literacy achievement across the whole age cohort.

By time point three, both the RR and TNI group made further progress but the RR group continued to outperform the TNI group and the differences between the results of book level and the writing vocabulary test were statistically significant. Although the RR group had higher mean scores on the external measures, the results were not statistically significant. One could argue that the tasks of the OSELA (Clay, 2013) were more sensitive to small changes in literacy learning or, as Slavin et al. (2011) argued, perhaps the tasks pick up changes because the tasks are close to the instruction they received. However, we agree with D'Agostino and Harmey's (2016) argument that the oral reading of books as per the running record and the writing of words are similar to classroom instruction in general.

The TNI group was found to have the highest mean scores on the BAS3 word reading test. However, this group was reduced from 22 to 10 as the lowest attaining children were now in RR. The TNI group results also demonstrated a large standard deviation (*SD* = 16.66) suggesting greater variation in the scores compared to the RR group.

Limitations

This study had a small sample size which limits the inferences that can be made from the study. Given that half of the TNI group entered the RR intervention mid-year, it was difficult to compare across groups at time 3. The gold standard for comparing across groups would be to randomly assign students to treatment or control conditions but this was not possible. Nevertheless, we suggest it was worthwhile to conduct some comparison of these naturally occurring groups to start to build preliminary data on the efficacy of RR in this context.

Another limitation is the short time span of this study. This was a relatively short amount of time within the context of a new implementation given an exploratory investigation. In future, it would be useful to conduct a longer, more comprehensive study with a larger sample and, if possible, random assignment.

Implications for Policy

In this section of the paper we discuss some implications for practice. To start, however, it is important to reflect on why preliminary studies like this are so important as a way to continue to build the evidence base for RR as it moves into different contexts. As stated at the beginning of the article, the Scottish Government has focused on achieving equity in education through the Scottish Attainment Challenge. This places a clear focus on accelerating learning through "targeted improvement activity in literacy" (Education Scotland, 2020). As a result of this policy and the targeted funding that was available to local authorities to invest in early intervention, RR was

introduced as one of the approaches to early intervention. As with any investment, it is incumbent on schools and authorities to demonstrate that the intervention is indeed reducing the gap.

This small-scale exploratory study has clearly demonstrated that, for the children in this study, accelerated learning was achieved and the gap, in terms of literacy attainment, was closed. This was evident in two ways. First, the children who were identified as the lowest attaining in literacy at the beginning of the year were, 5 months later, achieving age-expected levels as measured at time 2. Second, of the 22 children who were identified as experiencing difficulties but were not the lowest in September, 12 went on to enter RR later in the year, meaning ordinary classroom instruction did not help them catch up. Thus, the use of RR as one of the targeted improvement strategies by these local authorities was a good decision.

This study, though small, provides preliminary results of efficacy in the Scottish context. Bryk (2015) argues that starting small allows organizations to make sure the contextual structures work well together. What is needed now, however, is continued collection of school, local, and national level data at the time of intervention and over time as the children progress through school. Indeed, continued evaluation with more robust designs would continue to permit a nuanced examination of the variation of achievement in results. This allows the story of the efficacy of the intervention to be told and identification of unexpected results – both elements contribute to 'achieving better outcomes more reliably at scale' (Bryk, 2015, p.471).

Implications for Practice

In this section, we share two practical ways the results of the study have been shared at a local level. This was done via:

1. The Scottish Learning Festival
2. Poster

Scottish Learning Festival (SLF)

The SLF is the annual conference and exhibition for educational practitioners across Scotland. The results of the study were communicated through the use of a poster by the lead author at an exhibit at this event. The results were of interest to many teachers, Head Teachers and Literacy Leads from across Scotland. Due to the busy nature of the exhibition there was little time to communicate much in detail, however, it allowed for contact details to be passed on for anyone of interest so that meetings could be set up to share the results of the study and information about RR in more detail.

Poster

The poster used by the lead author was adapted to provide more detail on the study and the results (contact the lead author to request a copy of the poster). This was very useful for sharing with Head Teachers and Literacy Leads from different Local Authorities at meetings for those who had shown an interest in possibly having RR within their schools. Some of these meetings came about as a result of the SLF. It was much easier to communicate the results within a smaller more intimate setting as this provided a better environment for questions to be asked of the data and for explanations to be given. The graphs on the poster seemed to generate the most interest to these professionals and even more so when the book levels graph was explained in terms of reading ages. This tended to be met with astonishment. The fact that the data was obtained in Scotland has certainly made it more relevant and convincing to professionals within Scotland who are looking for ways to spend PEF money with a view to closing the attainment gap. RR meets the criteria for PEF as it is clear to see the impact that the intervention has on attainment, making it a worthy investment for Head Teachers. Many of the Head Teacher's and Literacy Leads felt that RR could provide the solution to ensuring that PEF is spent with a view to achieving equity in education.

They were excited that RR would not only achieve equity in education but that the results suggest long-lasting impact for the children who are targeted and clearly shows that the poverty related attainment gap can be closed.

Conclusion

All children included in the study made gains in their literacy levels, though the RR intervention yielded more positive effects. The children who received the RR intervention reached age expected levels by the end of the intervention and they continued to make gains greater than those of the children who did not receive RR. It would seem that the three authorities who decided to invest in RR within Scotland made a good decision as the short term effects of RR can be seen clearly thorough the results of this study. Whether the children in this study continue to make gains in the long term remains to be found. This study has highlighted the gains made by the RR group in comparison to a group of children who did not receive RR. Within the Scottish context, and in general, there are a lack of comparison studies in this area and what is now needed is further research implemented over a longer period of time and perhaps the initiation of gathering and studying longitudinal data. This would provide a stronger, evidence-based conclusion of the importance and efficacy of the decision to implement RR. This study has highlighted the importance of early intervention. It has proved RR to be an effective way of helping children to reach age expected levels in both reading and writing. The rigorous assessment format within RR allows for attainment to be tracked and therefor impact to be seen clearly. With all schools working to close the attainment gap, it seems that RR is an investment worth making and it can help schools meet the demands set forth by PEF and the goal of equity in education for all our students.

References

Bodman, S. (2019). *10 Years On: The Impact of Reading Recovery [online]* Available from: https://www.teachwire.net/news/10-years-on-the-impact-of-reading-recovery [Accessed 15 April 2019].

Burroughs-Lange, S. & Douëtil, J. (2007). Literacy Progress of Young Children from Poor Urban Settings: A Reading Recovery Comparison Study. *Literacy Teaching and Learning, 12,* pp.19-46.

Bryk, A. S. (2015). 2014 AERA Distinguished Lecture: Accelerating How We Learn to Improve. *Educational Researcher*, *44*(9), 467–477. https://doi.org/10.3102/0013189X15621543

Chapman, J.W. &Tunmer, W.E. (2019). Reading Recovery's unrecovered learners: Characteristics and issues. *Review of Education, 7*, pp.237 -265.

Clay, M.M. (1991). *Reading Recovery Surprises*. In D.E. DeFord, C.A. Lyons and G.S. Pinnell (Eds.), *Bridges to Literacy: Learning from Reading Recovery.* Portsmouth, NH: Heinemann.

Clay, M.M. (2013). *An Observation Survey of Early Literacy Achievement.* 3rd ed. Auckland: Global Education Systems Ltd.

Clay, M.M (2015). *Becoming Literate: The Construction of Inner Control.* Auckland: Global Education Systems Ltd.

Clay, M.M. (2016). *Literacy Lessons Designed for Individuals.* Auckland: Global Education Systems Ltd.

D'Agostino, J.V. & Harmey, S.J. (2016) An International Meta-Analysis of Reading Recovery. *Journal of Education for Students Placed at Risk (JESPAR), 21*, pp. 29-46.

Douëtil, J. (2005) Change and Opportunities for Learning. *The Journal of Reading Recovery Special Collections*, *1*, pp.76-79.

Education Scotland (2017). *Getting it Right for Every Child (GIRFEC)* [online] Available from: https://education.gov.scot/scottish-education-system/policy-for-scottish-education/policy-drivers/Getting%20It%20Right%20For%20Every%20Child%20(GIRFEC) [Accessed 11 March 2019].

Education Scotland (2019). *What is Curriculum for Excellence?* [online] Available from: https://education.gov.scot/scottish-education-system/policy-for-scottish-education/policy-drivers/cfe-(building-from-the-statement-appendix-incl-btc1-5)/What%20is%20Curriculum%20for%20Excellence [Accessed 26 February 2019].

Education Scotland (2020). Scottish Attainment Challenge. Available from: https://education.gov.scot/improvement/learning-resources/scottish-attainment-challenge/

Elliott, C.D. and Smith, P. (2011). *British Ability Scales 3: Word Reading Test.* London, UK: GL Education Group.

Foster, H (2007). *Single Word Reading Test 6-16: SWRT 6-16.* Brentford, UK: GL Education Group.

GL Education (2019). *Single Word Reading Test* [online] Available from: https://gl-education.com/products/single-word-reading-test-swrt/ [Accessed 30 May 2019].

Hurry, J. and Fridkin, L. (2018) *The impact of Reading Recovery ten years after intervention: A report for the KPMG Foundation.* UCL Institute of Education.

Lyons, C.A. (2003). *Teaching Struggling Readers: How to Use Brain-based Research to Maximise Learning.* Portsmouth, NH: Heinemann.

May, H., Sirinides, P., Gray, A.& Goldsworthy, H. (2016). (2016) Reading Recovery: An Evaluation of the Four-Year i3 Scale-Up . CPRE Research Reports.

Ogier, M.E. (2002). *Ogier's Reading Research. How to Make Research More Approachable.* Balliere Tindall.

Robson, C. & McCartan, K. (2016). *Real World Research (Fourth Edition).* West Sussex: John Wiley and Sons Ltd.

Schildkamp, K. (2019) Data-based decision-making for school improvement: Research insights and gaps. *Educational Research*, 6, 257-273, DOI: 10.1080/00131881.2019.1625716

Schmitt, M.C., Askew, B.J., Fountas, I.C., Lyons, C.A., Pinnell, G.S. (2005) *Changing Futures: The Influence of Reading Recovery in the United States.* Worthington, OH: Reading Recovery Council of North America.

Schwartz, R. M. (2018). Reading Recovery: How Do We Rank? *Journal of Reading Recovery*, *17*(2), 61-65.

Scottish Education Council (2018). *Pupil Equity Funding (SEC-02(02)).* Scottish Government Learning Directorate.

Scottish Government (2017) *Scottish Index of Multiple Deprivation* [online] Available from: http://www.gov.scot/collections/scottish-index-of-multiple-deprivation [Accessed 26 Feb 2019].

Scottish Government (2018a). *Pupil Attainment: Closing the Gap.* [online] Available from: https://beta.gov.scot/policies/schools/pupil-attainment/ [Accessed 26 Feb 2019].

Scottish Government (2018b) *Achievement of Curriculum for Excellence (CfE) Levels 2017/18* [online] Available from: https://www.gov.scot/publications/achievement-curriculum-excellence-cfe-levels-2017-18/pages/3/ [Accessed 26 Feb 2019].

Slavin, R.E., Lake, C., Davis, S. & Madden, N.A. (2011) Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review,* 6, pp.1-26.

Swinson, J. (2013) British Ability Scales 3, *Educational Psychology in Practice*, 29, pp.434-435.

What Works Clearinghouse, U.S. Department of Education, Institute of Education Sciences (2013). *Beginning reading intervention report: Reading Recovery.* Retrieved from https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_readrecovery_071613.pdf

Table 1. *Participants*

School Codes	SIMD* Decile	Total Number of Pupil Participants	Number of RR Participants (Number eligible for PEF**)	Number of TNI Participants (Number eligible for PEF**)	Number of TNI pupils who became part of cohort 2
B	2	4	3 (3)	1 (1)	0
C	10	5	3 (0)	2 (0)	0
G	5	5	2 (0)	3 (0)	2
H	2	9	2 (2)	7 (3)	3
I	9	8	6 (4)	3 (0)	2
K	1	3	2 (2)	1 (1)	1
L	10	5	2 (0)	3 (0)	3
N	2	6	4 (4)	2 (2)	1
Totals:		**46**	**24 (15)**	**22 (7)**	**12**

**SIMD – Scottish Index of Multiple Deprivation*

***PEF – Pupil Equity Funding*

Table 2. *Means and standard deviations of RR group on all assessments*

	N	Mean	Std. Deviation	Std. Error Mean
Book Level Time 1	24	1.83	2.46	.50
Book Level Time 2	24	15.67	3.42	.70
Book Level Time 3	24	17.00	4.66	.95
WV* Time 1	24	6.83	4.31	.88
WV Time 2	24	32.83	11.54	2.36
WV Time 3	24	35.00	8.88	1.81
B.A.S.3 Time 1	24	5.21	4.03	.82
B.A.S.3 Time 2	24	27.25	10.94	2.23
B.A.S.3 Time 3	24	30.88	10.97	2.24

**Writing Vocabulary*

Table 3. *Means and standard deviations of RR and TNI groups on initial assessments*

	Group	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
Book Level	RR	24	1.83	2.46	.50
	TNI	22	2.86	3.59	.77
Writing Vocabulary	RR	24	6.83	4.31	.88
	TNI	22	8.82	7.08	1.51
BAS 3	RR	24	5.21	4.03	.82
	TNI	22	7.91	7.35	1.57

Table 4. *Means and standard deviations of RR and TNI groups at Time 2 and Time 3*

	Group	*N*	*Mean*	*Std. Deviation*	*Std. Error Mean*
Book Level Time 2	RR	24	15.67	3.42	.70
	TNI	22	8.55	7.20	1.54
Writing Vocabulary Time 2	RR	24	32.83	11.54	2.36
	TNI	22	21	11.17	2.38
BAS 3 Time 2	RR	24	27.25	10.94	2.23
	TNI	22	18.00	13.63	2.91
Book Level Time 3	RR	24	17.00	4.66	.95
	TNI	22	12.68	6.58	1.40
Writing Vocabulary Time 3	RR	24	35	8.88	1.81
	TNI	22	27.14	14.05	2.99
BAS Time 3	RR	24	30.88	10.97	2.24
	TNI	22	25.00	14.38	3.07
Single Word Reading Test Time 3	RR	24	17.54	6.56	1.34
	TNI	22	13.55	8.18	3.27

Table 5. *Means and standard deviations of the RR, RR2 and TNI groups at time 3.*

		N	Mean	Std. Deviation	Std. Error
Book Level	TNI-	10	14.70	8.60	2.72
	RR	24	17.00	4.66	.95
	RR2	12	11.00	3.91	1.13
	Total	46	14.93	6.01	.89
Writing Vocabulary	TNI-	10	30.30	14.51	4.59
	RR	24	35.00	8.88	1.81
	RR2	12	24.50	13.70	3.96
	Total	46	31.24	12.17	1.79
BAS3	TNI-	10	31.40	16.66	5.27
	RR	24	30.88	10.97	2.24
	RR2	12	19.67	9.99	2.88
	Total	46	28.07	12.92	1.90
SWRT	TNI-	10	16.20	10.33	3.27
	RR	24	17.54	6.56	1.34
	RR2	12	11.33	5.35	1.54
	Total	46	15.63	7.57	1.12

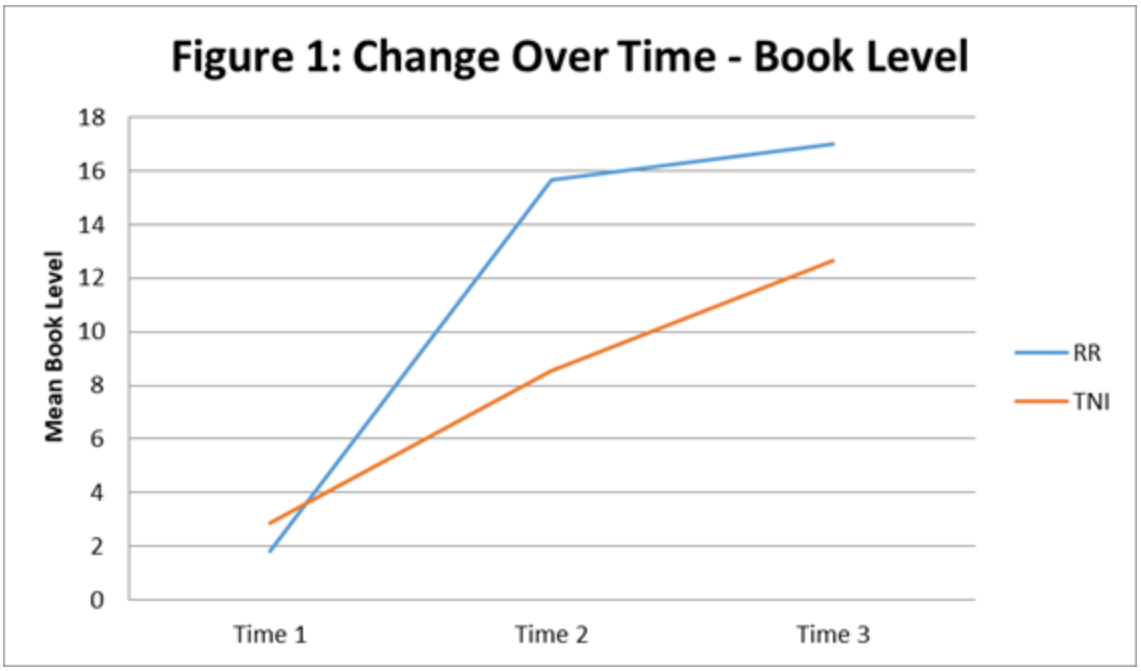
Figure 1: Change Over Time - Book Level
Mean Book Level
18
16
14
12
10
8
6
4
2
0
Time 1
Time 2
Time 3
RR
TNI

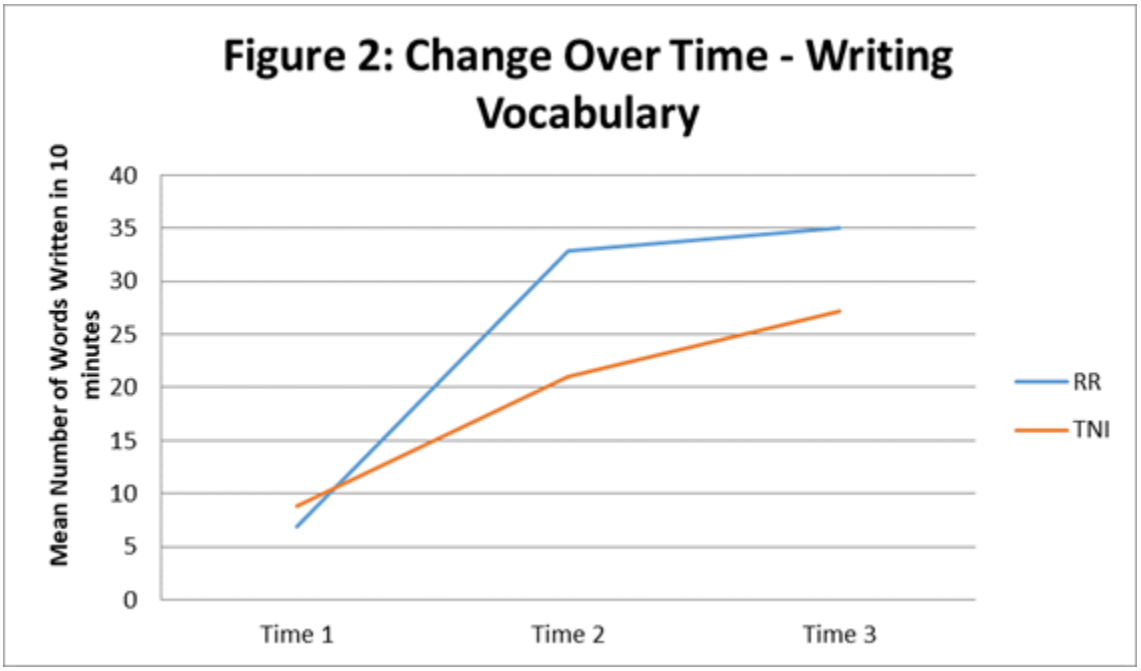
Figure 2: Change Over Time - Writing Vocabulary
Mean Number of Words Written in 10 minutes
40
35
30
25
20
15
10
5
0
Time 1
Time 2
Time 3
RR
TNI

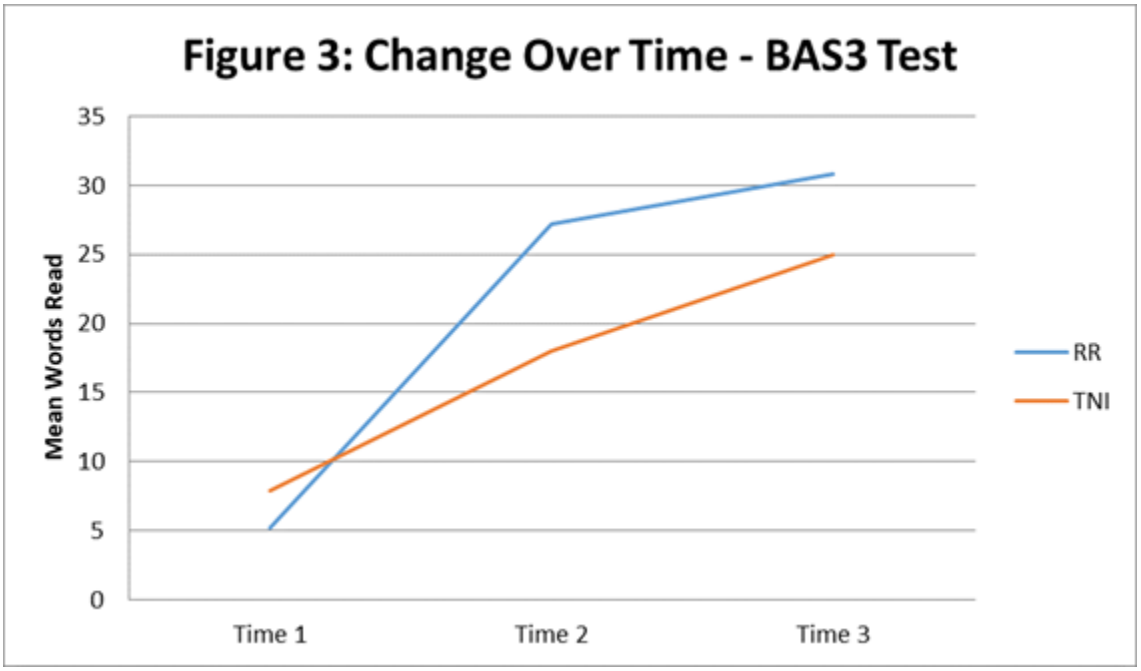
Figure 3: Change Over Time - BAS3 Test
Mean Words Read
35
30
25
20
15
10
5
0
Time 1
Time 2
Time 3
RR
TNI