An innovative classroom reading intervention for Year 2 and 3 pupils who are struggling to learn to read:

Evaluating the Integrated Group Reading Programme

EXECUTIVE SUMMARY AND PROJECT REPORT

MAY 2018

BRAHM NORWICH
GEORGE KOUTSOURIS
ALEXEY BESSUDNOV
UNIVERSITY OF EXETER
GRADUATE SCHOOL OF EDUCATION

Contents

1. Executive summary

The Integrated Group Reading (IGR) programme

The Integrated Group Reading (IGR) programme is a targeted teaching intervention for Year 2 and 3 pupils who are delayed in reading and is taught by class teachers in small groups during the existing small group organisation of lessons (Guided Reading or other form of group reading). It is part of a class-wide model, with all pupils being in groups receiving teacher attention over a period of a week, supported by a teaching assistant (TA). The term 'Integrated' refers not only to the inclusive aspect of the class-based organisation that enables pupils identified for targeted support access to teacher expertise alongside their peers, but also to the integration of several discrete professional and research-based approaches to literacy teaching underpinning its methodology.

The IGR programme is a response to three ongoing issues: firstly, almost 20% of children in English Primary Schools on entering Key Stage 2 (KS2) are delayed or non-starting readers (Department for Education (DfE), 2017), and analysis of the DfE phonics test in 2016 indicates that around 10% did not reach the nationally set threshold level at the end of Year 2. Secondly, there is an ongoing debate about the primacy of certain phonics approaches over others (synthetic versus analytic) but the research evidence is inconclusive (Henbest & Apel, 2017). Since the Rose (2006) report, English policy has favoured the synthetic phonics approach. However, the fact that 10 per cent of pupils taught using synthetic phonics still experience difficulties in reading suggests that other approaches should be tried for these children. Thirdly, current practice is to provide 'Quality First' teaching that is meant to be differentiated, but might not be differentiated enough for pupils struggling to learn. So tailored teaching for those not progressing at the expected rate with targeted or specialist teaching is often offered in the form of withdrawal sessions with people other than the class teacher (e.g. TAs).

This has two potential implications: i) it can create a 'separation' effect (Education Endowment Foundation (EEF), 2015) by limiting the opportunities of these pupils for quality time with the class teacher and peer interactions; and ii) it can mean learning time lost – for instance, it has been found that children who had immediate access to additional support rather than waiting to fail, had improved reading outcomes at the end of Year 1 (Al Otaiba et al, 2014). The IGR programme addressed all three ongoing issues, by introducing tailored targeted teaching (tier 2) in the 'Quality First' teaching setting, thus enabling the teacher to deliver a multi-perspective intervention (including phonics) to struggling pupils during a whole-class session that was literacy-related for all.

The IGR programme involves appealing books and games tied into an incremental progression, which are ready for busy teachers to use. The learning through books and games is story-focused to support the enjoyment of reading and encourage deep pupil engagement with the text.

The IGR programme trial

IGR was trialled by the Graduate School of Education of the University of Exeter with Year 2 and 3 pupils in 34 English schools in four varied local authority areas across two years (2015-2017). The programme was delivered four times a week for 30 minutes over two terms during whole-class sessions as part of the usual group reading organisation (typically the class is organised into four to six groups). The teacher taught the IGR group (comprised of four pupils identified by their teacher as in need of literacy support) twice a week and introduced a new book at each session. TAs worked twice a week with the group in-between the teacher sessions for consolidation. During teacher-led IGR, the rest of the class (organised in reading groups) worked independently or with a TA on various reading-related activities. Teachers would work with other groups once a week on a carousel basis when not teaching the IGR group. So, teachers and TAs had discrete yet interconnected roles, with

the teachers keeping the main role. IGR was designed to be part of the usual group reading schedule, allowing teachers to organise their group reading rota in a more structured way for all pupils.

The IGR programme was evaluated using a mixed methodological design, involving i. a clustered randomised control trial with the comparison group in control schools (phase 1) and the control schools using IGR in phase 2, ii. a process evaluation of implementation and teachers' and pupils' experiences of using the programme, and iii. teaching case studies where the quality of IGR teaching was and was not related to the extent of reading gains.

Key conclusions

Experimental evaluation

Participating children in schools using IGR in phase 1 and phase 2 made the same degree of progress in reading accuracy/ comprehension, compared to similarly struggling children in control schools who were mainly using phonics approaches (no statistically significant differences). The mean reading progress in intervention and control groups was equivalent to 11.5 months over the 7 months in phases 1 and 2 using the mean of two measures of reading, often seen as 'modest impact'. In phase 2 one reading test ((accuracy and comprehension) showed a gain of 14 months, ('useful impact') not shown by the other test (word reading). There were also no statistically significant changes for reading and school attitude in either the treatment or control group. This suggests that our initial hypothesis that IGR would improve reading gains and attitudes for the IGR group compared to the control group was not supported by the findings.

There were no statistically significant differences between boys/girls and Year 2/3 pupils in their responses in the IGR and control classes. Some analyses showed that pupils having English as an Additional Language (EAL) and being identified for Pupil Premium made significantly greater gains with IGR, but these findings were not replicated across phases, measures or levels of significance.

In Phase 1 there was no statistically significant difference in gains between treatment and control classes for non-IGR children. This confirms our initial hypothesis that IGR in the classroom would not have any negative effect on the classroom pupils not having the intervention. In Phase 2, non-IGR children showed somewhat better progress on the Hodder standardised scale in the treatment classes compared to the control classes (d = 0.2). This effect was statistically significant, but this is interpreted as probably due to the high baseline scores for the girls in the control group and possibly because of a measurement error.

For teachers using IGR, their self-efficacy in teaching reading through a self-report measure improved significantly in both phases. Control teachers did not complete this measure, so this change is hard to interpret.

Process evaluation

Overall, participants were enthusiastic about the intervention, the project materials, and accompanying support. Teacher-reported outcomes for IGR pupils included increased confidence, motivation and interest in reading, and improved reading, oral language and social skills. Some teachers were concerned that these gains had not yet transferred outside of the IGR group setting. Most pupils were not worried about being seen in a low attainment group, and did not see IGR as an intervention, but as an exciting classroom activity. Other class pupils were often very interested in the IGR resources, especially the games.

IGR was used with varied fidelity, and many teachers had limited understanding of the theory underpinning the programme, which could partly indicate a training limitation. In phase 1, this resulted in the programme support team having to produce a table with acceptable and unacceptable variations to programme implementation to advise teachers accordingly. This reflected individual variation in the way the programme was used, and that fidelity was operating along a continuum. Some departures from the suggested methodology were seen in some cases to be justifiable (such as slowing down the pace of the programme in response to pupils' needs), whereas others were less acceptable (for instance, delivering all programme sessions in withdrawal sessions).

In addition, control schools did not just continue with typical teaching; teachers recognised that control pupils had significant additional needs, so they also had a great deal of additional, mainly phonics-based teaching input, making what was being compared to the experimental evaluation varied and complex.

Case studies

Data from two different teachers showed that when high reading gains followed high IGR teaching fidelity, several supportive factors were identified, e.g. teacher and pupil enthusiasm, school leader and adviser involvement, teacher understanding the theory and rationale of IGR and the IGR model fitting the pre-existing reading organisational arrangements. When low gains were followed by low fidelity in the cases of two other teachers, the above factors were not identified.

In teacher cases where low or no reading gains followed medium to high IGR fidelity, there was evidence of factors that were barriers to reading progress, such as a mechanical teaching approach that did not engage pupils, having a TA who could not manage the other groups during IGR teaching, and unsatisfactory teacher job-sharing arrangements. In the case of one teacher where quite high reading gains followed low IGR fidelity, there was evidence that the fidelity measure was affected by a change in teaching which did not affect the otherwise high quality IGR teaching.

Overall findings

The experimental evaluation indicates that the multi-perspective IGR approach that supports enjoyment of reading resulted in as much reading gain as the more phonics-oriented programmes used in control classes. The process evaluation and case studies illustrate further benefits and some challenges not found in the measured outcomes. This means that IGR might be considered by schools and teachers as an alternative to the current pattern of targeted interventions that involve more phonics-based programmes delivered by TAs.

The lack of negative effects on reading in non-IGR pupils is a noteworthy finding suggesting that using IGR in the classroom had no negative effects on other children's progress. This reinforces the use of IGR as an alternative to the usual model of offering additional support in withdrawal sessions often led by TAs.

The gains in teachers' perceived self-efficacy in teaching reading following the use of IGR point to possible confidence benefits in teaching reading.

While the process evaluation showed IGR strengths (e.g. teacher enthusiasm and enhanced pupil confidence) and limitations (e.g. insufficient understanding of the theory of IGR teaching), the teaching case studies nevertheless illustrated how useful reading gains depended on IGR teaching fidelity and other supportive factors.

The programme does not have considerable implementation costs apart from the one-off cost of materials; however, it has particular staffing demands (teachers and TAs), and it can be used more effectively in classrooms that have a regular TA.

Significance of the key findings

IGR adopts a multi-perspective approach that includes phonics, but also enables pupils to engage more deeply with text and allows for comprehension to emerge naturally. Following the Rose (2006) report, schools in England largely use phonics approaches to teach early reading, with explicit phonics teaching showing good results (Wyse & Goswami, 2008).

The experimental evaluation indicates that the multi-perspective IGR approach that supports enjoyment of reading resulted in as much reading gain as the more phonics oriented additional programmes used in control classes (as evidenced by the data collected from control schools). Since we had decided not to intervene with the teaching decisions in the control classes, the trial was in fact comparing IGR to a programme influenced by the dominance of the phonics approach and driven by the national curriculum and the assessment requirements. This means that IGR might be considered by schools and teachers as an alternative to the current pattern of targeted interventions that involves more phonics-based programmes delivered in most cases by TAs.

The IGR trial also found that IGR organisation made it possible for the teacher to deliver targeted (tier 2) teaching to pupils who are delayed in their reading in the regular class, without this having a negative impact on the rest of the class pupils. These findings are relevant to teachers, advisers and policy makers who are looking for more inclusive approaches for targeting pupils in Years 2 and 3 who are delayed in their reading.

Implications about how additional support is organised

The IGR organisation that enables teachers to offer targeted (tier 2) teaching in a 'Quality First' (tier 1) setting proved to be challenging but viable. This has implications about the way additional provision is organised for pupils identified as in need of targeted support. It particularly shows how it is practically possible for the teacher to take responsibility for the learning of all pupils, even by offering extra time to those most needing it, without hindering the learning of the rest of the class.

This model for the organisation of additional support (group organisation, coordinated teacher-TA collaboration and well-prepared materials) could be extended beyond reading to other areas of learning, e.g. aspects of mathematics or science. Future research and development might explore this approach to inclusive targeted support beyond the teaching of reading.

Areas for future research and development

As a first step after completing the study, there is scope to explore how and whether IGR is being used by teachers in participating schools with the restrictions of the RCT protocol removed. It is likely that teachers would use the programme materials mainly in a loose way, but there is still value in exploring the reasons behind their decisions.

Future studies could also be designed with greater focus on teacher professional learning about the principles of IGR and more focussed training and coaching of IGR-related teaching skills.

With regards to programme development, the IGR programme developer, building on teacher feedback, has designed a synthetic phonics game that could be added readily to future versions of IGR (the version trialled here had an analytic phonics component). This game is story-specific and fits with the existing programme principles and materials.

2. Brief description of existing practice and justification for an alternative approach

Almost 20% of children in English Primary Schools on entering Key Stage 2 (KS2) do so as delayed or non-starting readers (Department for Education – DfE, 2017), and analysis of the DfE phonics test in 2016 indicates that around 10% did not reach the nationally set threshold level at the end of Year 2 (DfE, 2016). This persistent challenge can be attributed to several factors including the opaque nature of English orthography (Wyse & Goswami, 2008). Though impressive attempts are made in Key Stage 1 (KS1) to ensure that all children acquire and can deploy the phonic knowledge that they will need as a basis for both encoding for writing and decoding for reading, national statistics only show modest gains between 2006-13 (DfE, 2013), although gains are more significant after 2013 (DfE, 2016). This leaves some children unable to make sufficient progress in reading to be able to benefit in a full way from an increasingly lively and diverse KS2 curriculum in the context of classroom 'Quality First' teaching.

The Response to Intervention (RTI) model has often been used to describe how teaching is geared towards the needs of pupils who are struggling to read (Fien et al, 2011; Griffiths & Stuart, 2013). The RTI model distinguishes between what is offered to all (tier 1) or 'Quality First' in the UK, and to some (tier 2) or to a few (tier 3). We will use the term 'tier' in this report and not the alternative term 'wave'.

'Quality First' teaching can be differentiated to address a range of needs. However, current practice is to provide 'Quality First' teaching that is not differentiated enough for pupils struggling to learn (for practical and time reasons, lack of skill or resources), and then provide more tailored teaching for those not progressing at the expected rate in higher and differentiated tiers (2/ 3); this is often offered as pull-out sessions with people other than the class teacher (e.g. teaching assistants – TAs). It has been suggested though that this can create a 'separation' effect (EEF, 2015), as it limits the opportunities of these pupils for quality time with the class teacher (whose time and attention they are the most in need of) and peer interactions. Also, without sufficient differentiation, the time children who struggle to learn spend in insufficiently differentiated 'Quality First' settings has been seen as learning time lost (Al Otaiba et al, 2014).

An additional matter relevant to tier 2/ 3 remedial programmes is the approach to teaching reading. This reflects the ongoing debate about the primacy of certain phonics approaches over others (synthetic versus analytic). Although since the Rose (2006) Report English policy favours the former, the research evidence is inconclusive (Henbest & Apel, 2017). Yet, there are literacy programmes which are popular with schools in England that rely heavily on a single synthetic phonics approach.

3. Short overview of IGR

The Integrated Group Reading (IGR) programme has been designed in response to these issues. The programme is a tier 2 intervention targeting Year 2 and 3 pupils who are delayed in reading and is taught by class teachers in small groups during the existing small group organisation of lessons (Guided or other form of Group Reading). It is part of a class-wide model, with all pupils being in groups receiving teacher attention over a period of a week, supported by a teaching assistant. The term 'Integrated' refers not only to the inclusive aspect of the class-based organisation that enables pupils identified for tier 2 support access to teacher expertise alongside their peers but also, significantly, to the integration of several discrete professional and research-based approaches to literacy education underpinning and justifying its methodology.

IGR is original in two particular ways: Firstly, it introduces a tier 2 targeted intervention into the 'Quality First' setting. A literature review that was conducted by the authors to explore the nature of

additional support for children struggling to learn to read found no systematic evaluations of programmes using similar delivery arrangements. It was found that the people responsible for teaching wave/tier 2 interventions evaluated in the UK were mostly teaching assistants in pull-out sessions (e.g. See et al, 2015; Clarke et al, 2010; Duff et al, 2008; Hatcher et al, 2006). This could be to some extent indicative of how additional support is often organised in UK primary schools. Although there is a risk of over-generalising to situations where appropriate support is provided by teaching assistants, pupils who are struggling to learn to read might be seen to need their teacher's attention more than other pupils.

Secondly, given the dominance of synthetic phonics teaching in England following the Rose (2006) report, IGR represents a unique blend of approaches to teaching reading to early readers who are struggling. The programme adopts a multi-perspective approach to reading, integrating (analytic) phonics elements, story-telling for oral language development, word games, Paired Reading approaches, and elements from Reading Recovery, as discussed below. By doing so, IGR moves away from usual Guided Reading protocols aiming to make pupils independent readers and focuses on simplicity to support the enjoyment of reading from which understanding is expected to follow naturally.

Phonics teaching

The IGR programme has currently a mainly analytic phonics component based on onset and rime. Substantial research shows that phonics teaching is not only the most examined teaching approach, but the one with the greatest efficacy for reading and spelling gains (Department for Children, Schools and Families – DCSF, 2008; Galuschka et al, 2014; National Early Literacy Panel – NELP, 2008; National Reading Panel – NRP, 2000), however the efficacy of different approaches is still debated (Henbest & Apel, 2017). Several studies consistently report no difference between synthetic and analytic phonics (Comaskey et al, 2009; Di Stasio et al, 2012; Ehri et al, 2001; Kyle et al, 2013), yet there are few studies supporting that there is a difference (synthetic over analytic phonics: Christensen & Bowey, 2005; Johnston & Watson, 2004; analytic over synthetic: Walton et al, 2001).

Oral language skills

In addition, early progress in reading has been shown to depend on children's oral language skills (Bowyer-Crane et al, 2008; Clarke et al, 2010; Muter et al, 2004; Nation & Snowling, 2004) – e.g. Clarke et al (2010) found that, between three treatment groups, an oral language training treatment group made better gains in reading comprehension between the end of the intervention and an eleven-month follow-up. They concluded that difficulties in reading comprehension partly reflect underlying oral language weaknesses calling for suitable teaching. Bowyer-Crane et al (2008) also found that oral language training programmes can improve vocabulary and grammatical skills.

IGR has also a strong story-telling aspect aiming to enable the development of children's oral skills. Story-telling has particularly been explored as a way to promote language and literacy development, especially in the early years, with various study designs (Isbell et al, 2004; Peck, 1989; Sulzby, 1985; Valencia & Sulzby, 1991).

Word games

Research shows that word games can support the reading skills and engagement of children who struggle to learn to read (Raffaele Mendez et al, 2016; Jasmine & Schiesl, 2009; Charlton et al, 2005; Blachowicz & Fisher, 2004; Maclean et al, 1987). For instance, Raffaele Mendez et al (2016) report significant pre- and post-test gains for a reading intervention named Reading by Design that involves

board games to practise sight words and word attack skills, and foster engagement. In addition, Charlton et al (2005) found that games can accelerate learning when they are combined with teacher instruction. The IGR programme routine includes four reading games, each with a distinct pedagogic role (overlearning consolidation at the level of word and sentence, phonics practice, and advance organisation at the level of unfamiliar vocabulary).

Elements associated with Reading Recovery

The importance of detailed responses to reading in small groups as well as one-to-one contexts, and the monitoring of reading over time are elements that can be associated with Reading Recovery (Clay, 1994; Doyle, 2013; What Works Clearinghouse, 2013). Regarding the monitoring of reading, Clay (2001) wrote: 'The teacher would foster and support active constructive problem-solving, self-monitoring, and self-correction from the first lesson, helping learners to understand that they must take over the expansion of their own competencies. To do this, the teacher would focus on process variables (how to get and use information) rather than on mere correctness and habitual responses and would temporarily value responses that were partially correct for whatever they contributed toward correctness' (p. 652, cited in Doyle, 2013). IGR operates by engaging children at the cognitive level through word and phrase game-playing dynamics and group reading, encouraging collaborative problem solving, and enabling teachers to monitor progress closely.

Elements associated with Paired (Shared) Reading

A key element of the IGR approach to small group reading is its distinctive individual/choral collaborative approach to reading and problem solving with the teacher as leader/participant. Insofar as further relevant school-based research is concerned, the practice of Shared Reading can be said to be related to the strategic use of choral reading for the establishment of pace, fluency, and the modelling of speech-related cadences in the IGR approach. The whole class shared reading practice of 'Chiming In', for instance, which is a scaffolded, fluency-supporting approach for children who may have received less modelling of the act of reading, and who may have poor self-image as readers. For example, Gerde & Powell's (2009) study explores and analyses teacher book reading practices and their relation to children's receptive vocabulary, and found that pre-schoolers made greater gains in receptive vocabulary when their teachers used more book-focused utterances.

Experience from Paired Reading also suggests that this can be a valuable approach (Brooks, 2016; Miller et al, 2010; Topping, 2014; Topping et al, 2011; Topping & Lindsay, 1992). Topping (2014) describes the Paired Reading (PR) method as 'a form of supported oral reading which enables students to access and comprehend texts somewhat above their independent readability level […] This structured support used with high motivation texts offers dysfluent readers a flow experience, which is likely to impact on their reading style and socio-emotional aspects of the reading process' (p. 59). Topping et al (2011) report significant gains from a randomised control trial exploring Paired Reading in reading and self-esteem measures. Miller et al (2010) also found significant gains in self-esteem using a pre-and post-test design.

The delivery of structured, teacher-led supported oral reading in IGR groups of four children contrasts with intervention programmes of similar intensity that mainly have a one-to-one focus (Brooks, 2016). This aspect of the design of the IGR programme can be seen in the context of the Elbaum et al (2000) meta-analysis finding that comparing one-to-one with small-group supplemental instruction showed no advantage for the one-to-one programmes.

Programme materials

IGR provides an extensive set of learning materials, thus enabling teachers to work deeply and systematically with children struggling to learn to read. It has been designed to replace group reading work in the classroom until children have become confident early readers.

The IGR approach assumes that children have already been exposed to systematic phonics teaching; at the same time, it systematically supports children who may be struggling to translate this into the ability to read short running texts (stories) with accuracy, fluency, comprehension and confidence.

The success of IGR rests to a large extent on its appeal to the children themselves; the integrated, story-specific games allow them to relax and enjoy both overlearning and preparation for learning. Moreover, the short, gently staged story materials have been written to meet the requirements of pupils who may still need more in the way of early language experience that comes from story and rhyme than may otherwise be available to them as later-learning beginner readers.

The programme was using a range of specially written reading books with simple illustrations and accompanying story-specific games. The vocabulary content and gradient of the reading books and games is based on the IGR classification of English orthography (see: 'Mapping Standard Measures to the Acquisition of Phonics and Whole Words in Narrative Texts', Appendix 1) and a systematic, incremental progression through this. They are also written with the narrative requirements of later-learning readers in mind and are deliberately short so that one story can be completed in each lesson for best learning outcomes i.e. they are an integral part of the methodology and IGR model. There were enough IGR reading packs for teachers to be able to deliver the reading programme twice a week for 26 weeks (52 packs, example in Appendix 2).

At the time of the evaluation, the IGR programme materials began at red/yellow readability level (reading age equivalency 5.07 yrs.) and progressed through to turquoise readability level (reading age equivalency 7.01 – 7.04 yrs.). In some cases, additional materials were needed to cover the range of pupils' reading abilities (which was very varied across different LEAs). On the whole, it was found that more additional materials were needed at both the lowest and highest readability levels (vertically) but also within each readability level for pupils who reached a reading plateau (horizontally). Some readability levels seemed to need more materials than others, such as the yellow/blue and blue readability levels (lower middle), as some pupils were struggling to progress beyond these levels. Additional resources were produced using either original material or books readily available to the teachers.

IGR programme pilot

The programme materials and methodology were trialled initially on a twice-weekly basis over a one-term period by 7 teachers in four schools in Cornwall (2014-15). The aim of this pilot (funded by the University of Exeter) was to trial the IGR programme, materials and support system with Year 2 and 3 teachers and TAs in preparation for the implementation of the controlled evaluation. An earlier three-year period of developmental work (2010-13) on a small group once-weekly basis with 14 reading delayed Year 3 children in two schools in Devon and Plymouth reported useful gains for both reading accuracy and comprehension.

4. Brief overview of the evaluation project

IGR was trialled with Year 2 and 3 pupils in English primary schools who were delayed in their reading across 2 years: November 2015-May 2016 (phase 1) and October 2016-April 2017 (phase 2).

Details about RCT registration and the IGR team can be found in Appendix 3.

Figure 1. The Integrated Group Reading project's timeline

<table>
<tr><td>Time 1
Sept 15</td><td>Phase 1
Sept 15-July 16</td><td>Time 2
July 16</td><td>Phase 2 period
Sept 16- May 17</td><td>Time 4
May 17</td></tr>
<tr><td>Time 1</td><td>Intervention schools</td><td>Time 2</td><td></td><td>Time 4</td></tr>
<tr><td rowspan="4">PHASE 1
1. Assess identified pupils
2. Class reading assessment</td><td>October training: Year 2 & 3: IGR programme</td><td rowspan="4">1. Assess identified pupils
2. Class reading assessment
PHASE 2 T1</td><td>Year 3 not use IGR
Year 2 might use IGR in Year 3</td><td rowspan="4">1. Assess identified pupils
2. Class reading assessment
PHASE 2 T2</td></tr>
<tr><td>Typical teaching (control) schools Year 2 & 3: typical teaching</td><td>Time 3
(Sep 16)</td></tr>
<tr><td></td><td>Control schools use IGR. Some classes were control classes in phase 1; others not</td></tr>
<tr><td></td><td>Process evaluation 1</td><td></td><td>Process evaluation 2</td></tr>
</table>

The study explored the following questions:

- What were the immediate and long-term (9 months after the end of phase 1 implementation) effects of the IGR programme – in reading accuracy and comprehension, reading attitude and overall attitude to school – after its first and second year of implementation (phase 1 & 2) with Year 2 and 3 children identified as most struggling in reading, compared to similar children experiencing usual teaching?
- What were the immediate and long term reading gains for the rest of the classroom children (not receiving IGR) in the Year 2 and 3 classes that used the IGR programme with those most struggling in reading, compared to similar children in classes experiencing usual teaching?
- About the context and process of the IGR programme:
    a. What were the school and class contexts in which the IGR programme was used?
    b. How was it implemented as part of the whole class organisation model in the intervention schools?
    c. To what extent did teachers using the IGR methods have increased knowledge about teaching reading and increased self-efficacy to teach pupils struggling to learn to read?
    d. What were the perspectives and experiences of teachers/ pupils on using the IGR programme?
    e. How was reading taught in the control classrooms?

f. What was the fidelity of the programme implementation?

4.1. Implementation of IGR

IGR was designed to be delivered 4 times a week for 30 minutes as part of the usual Guided (or other Group) Reading session for all pupils. The teacher saw the IGR group twice a week and introduced a new book at each session. The TA worked with the group in-between the teacher sessions for consolidation. Teacher and TAs had very discrete yet interconnected roles, with the teacher keeping the main role (for example new books are never introduced by TAs). During teacher-led IGR, the rest of the classroom worked independently or with a TA on various reading-related activities (such as comprehension tasks, dictionary work, and computer literacy programmes). IGR was designed to be part of the usual Guided (or other Group) Reading classroom schedule, while allowing teachers to organise their Group Reading rota in a more structured and efficient way for all pupils. Teachers and TAs were encouraged to communicate regarding the pupils' reading progress daily, using built-in forms of communication (a Daily Record form).

Table 1 presents the 9 steps of the IGR intervention, shown separately for teachers and TAs.

Table 1: The IGR routine (for teachers/ TAs)

Teacher-led IGR Lesson:	**TA follow-up Session:**
Acknowledge TA work for previous book	Drawing and writing
Go Fish game for the previous book (sentence-based game)	Individual re-reading of yesterday's new book
New book introduction - the teacher tells some of the story and shows the pictures	Pelmanism* game with words from the story
Lotto game. Lotto helped pupils get used to the new book before reading it	
Choral and individual reading of the new book – also collaborative problem solving	
SWAP phonics game (based on the analytic phonics approach)	
**Pelmanism is card matching game involving memory of words*	

With the suggested IGR organisation, the teacher had to deliver IGR for 30 minutes twice a week. This meant that teachers had to re-think the organisation of their Guided Reading classroom carousels. For instance, where there were four Guided Reading groups in each class, teachers had to fit these four groups into the remaining three sessions (as there can be five sessions a week). One possibility was that, over a period of four weeks, children in the four Guided Reading groups would have the teacher taking them only three times, thus missing one session every four weeks. However, many teachers made it clear that this was unacceptable (due to school policies or parental concerns) and that all children needed to have the chance to see the teacher every week.

Teachers came up with a variety of solutions to this issue with the most common being delivering one of the teacher-led IGR sessions in the classroom but during a school assembly, so in a sense the IGR pupils were not pulled out of the class, but the rest of the pupils were not present. Other teachers used their SENCO time (as some were SENCOs) or asked their jobsharer to join them for one session to work with different groups. A few teachers were also reading with two groups on the same day (for 15 minutes with each or at different times in the day).

Some schools had issues with TA availability (relevant to funding and school-budget restrictions), which made IGR difficult to implement.

Costs were calculated using the original budget and a spreadsheet with all the programme-related expenses of the intervention (material production, main and follow-up trainings and support). Costings are based on Phase 1 and 2 spending figures per item, including both Nuffield and school paid elements. They also include an estimate for what was provided at no cost to the project in both phases by two Local Authorities. Costs per classroom include the one-off cost of materials that would not apply to subsequent years of programme use. Cost per IGR class was around £1,6500 (details in Appendix 4).

Initial and follow-up training

The training day for the phase 1 intervention teachers took place in Bristol in October 2015 and covered the IGR methodology and aspects of the classroom organisation. The training was revised following teacher responses and feedback from a survey sent to all attendants soon after the training, and the experience gained from the phase 1 implementation, and phase 2 teachers (of whom many were phase 1 control teachers) were trained in July 2016, also in Bristol.

The training was organised by the programme team who operated separately from the evaluation team. It offered support to the participating teachers in all aspects of IGR (IGR methodology, teaching strategies, programme materials, any type of programme-related concerns).

The teachers were also given a follow-up half-day training in their own local authority areas halfway through the intervention. In phase 1 these included explaining to teachers the principles behind the construction of new IGR materials for books readily available in their schools. Some teachers reported that this training was useful, but others felt that it was not relevant to them as they did not have the time nor the inclination to produce new materials. All teachers agreed that producing new materials would be time-consuming. So, when extra resources were needed the programme coordinator helped by preparing and supplying games and materials using, when possible, books readily available to schools. In response to this feedback, in phase 2, the follow-up training did not focus on material construction but recapped aspects of the IGR methodology.

Programme support

Programme team members visited all the participating schools at least once (and up to 3 times in Phase 2). This was also done in collaboration with Local Literacy Advisers and Education Improvement Officers who also visited schools in their areas at least once. They observed the programme's sessions and gave verbal and written feedback to both their teachers and the central team. The programme coordinator was in close contact with other programme team members including Literacy Advisers, school literacy leads and intervention teachers throughout the course of the year, although it took a long time in some instances to establish stable and effective communication networks. The quality of support developed over time as the people responsible for supporting the teachers became more experienced. This is evident by the very limited amount of programme-related issues reported by the phase 2 teachers (e.g. shortage of materials, lack of clear instructions) compared to the number of issues reported in phase 1.

In addition to the physical visits, the programme's website offered a great range of resources, information and supportive material that were regularly updated throughout the study. The

programme's website (http://www.integratedgroupreading.co.uk/) underwent three different phases of complete redesign to become more attractive, rich in content and user-friendly, since only a handful of teachers reported using the earlier versions.

Acceptable and less-acceptable variations to the programme

Halfway through phase 1 of the intervention (February 2016), as a consequence of both the programme and the evaluation arms' concern at the scale and complexity of departures from the methodology under evaluation, it was considered necessary to ask all teachers to bring their practice into better alignment with the intervention. To this end, a table with 'acceptable and non-acceptable' variations was produced and shared with the teachers (Appendix 5). An example of an 'acceptable' variation was using one book per week (instead of two) for groups that required more time, by spreading the lesson routine across two sessions. An 'unacceptable variation' would be to omit aspects of the programme routine, such as the phonics game at the end of the session. At the follow-up training which began in the second half of the intervention in phase 1, the programme team re-emphasised the importance to the project of full fidelity to the lesson methodology as well as explaining once again the underlying pedagogical rationale of the teaching approach. This table served also as guidance for the phase 2 implementation.

4.2. Evaluation methods overview

Trial design – phase 1

The IGR phase 1 programme evaluation had a mixed methodological design, involving:

- A clustered randomised control trial (clusters at the school level) with the comparison group in control schools on a waiting list to use the intervention.
- A process evaluation of implementation and teachers' and pupils' IGR experiences. This involved in-depth school level case studies, and a 2-weekly log to monitor the fidelity of implementation.

Trial design – phase 2

The phase 2 evaluation involved:

- A quasi-experimental study, in which the outcomes for the control schools in Phase 1 were compared with the outcomes in the same schools in phase 2 when they offered IGR.
- A process evaluation of implementation as in phase 1

Note that this design allows us to make two comparisons in this trial: In the first comparison, we compare the pupil outcomes in the randomised treatment and control schools in phase 1. In the second comparison, we compare the pupil outcomes in the control schools in phase 1 with pupils in phase 2 receiving IGR. Some children were control cases in phase 1 and IGR cases in phase 2, which has offered the advantage of allowing an internal comparison of the same children going from control to intervention conditions.

Participant selection – school recruitment

The project had the support and co-operation of Literacy Advisers in four local authority areas [in the South West (1), West Midlands (2) and Greater London (1)] who were actively involved in the recruitment process (and later in a supportive role). We originally aimed to recruit 40 schools with many schools expressing an interest in participating in the project (at one point almost 50) and

around 25 schools already had signed the project's memo of understanding in the summer term before the beginning of the project. However, in September 2015, when the programme's assessments were scheduled to begin, more schools decided to pull out from the study than we expected (including 5 who had signed the memo), despite having agreed earlier to participate in the project. A variety of reasons were offered, including: the Head teacher who signed up to the project had retired or changed, other research programme obligations, limited teacher availability, the cost of the materials, staffing and time requirements of the programme and assessments etc. In addition, the project's evaluation team had to arrange the school assessments within September 2015 to keep to the planned timetable. This also put pressure on some schools, with some interpreting the repeated attempts for communication (for signing documents, sending out ethics letters, arranging for various types of assessments, giving clarifications) as lack of coordination. There were also some Head Teachers who had signed the memo early in the summer but agreed to participate without fully reading the details contained in what they had signed. There was limited communication in some cases between the Head Teachers and teachers regarding the project's details and requirements. We also perhaps over-estimated the influence of Literacy Advisers in their local areas (they had initially approached schools) and our ability to finalise recruitment and arrange for the assessments in such a short period. In the limited time available we focused on recruiting more schools and managed to sign up 32 schools for the duration of the study.

In phase 2, 13 out of 16 schools continued in phase 2. The 3 schools that decided not to use the intervention (due to staffing and organisational difficulties) still agreed to participate in the remaining programme assessments (the assessments testing for long-term effects). Two more schools were then recruited from another South-West LEA (Figure 2).

Pupil identification

Both the IGR and control focus groups were identified by teachers following a standard procedure, using the same template and instructions adapted from Speece et al (2011). Speece et al (2011) had found that teacher rating is an accurate and efficient predictor of early reading difficulties, with the additional advantage that teachers have intimate knowledge of their pupils and that a teacher rating system is less time-consuming and more cost-effective compared to a standardised assessment.

The teachers were asked to identify four pupils who would benefit from literacy support and were given the questionnaire form and instructions. The selection was based on a teacher report scale that included reading attainment and attitude. The identification was done in two stages with the teachers being asked to identify up to 10 pupils who would benefit from literacy support, and then being asked to use a more refined version of the previous scale to decide on the final 4 pupils (the form in Appendix 6). The teachers determined the attainment levels by reference to class reading levels and not standardised scores.

The research team offered advice and support to make sure that the selected pupils would be able to access the IGR materials where adaptations were not possible (as for instance in a case of a pupil with severe learning difficulties, and another case where a school had a unit for pupils with hearing impairment). School leaders had the final word on this, as they had expert knowledge of the pupils' needs and learning style.

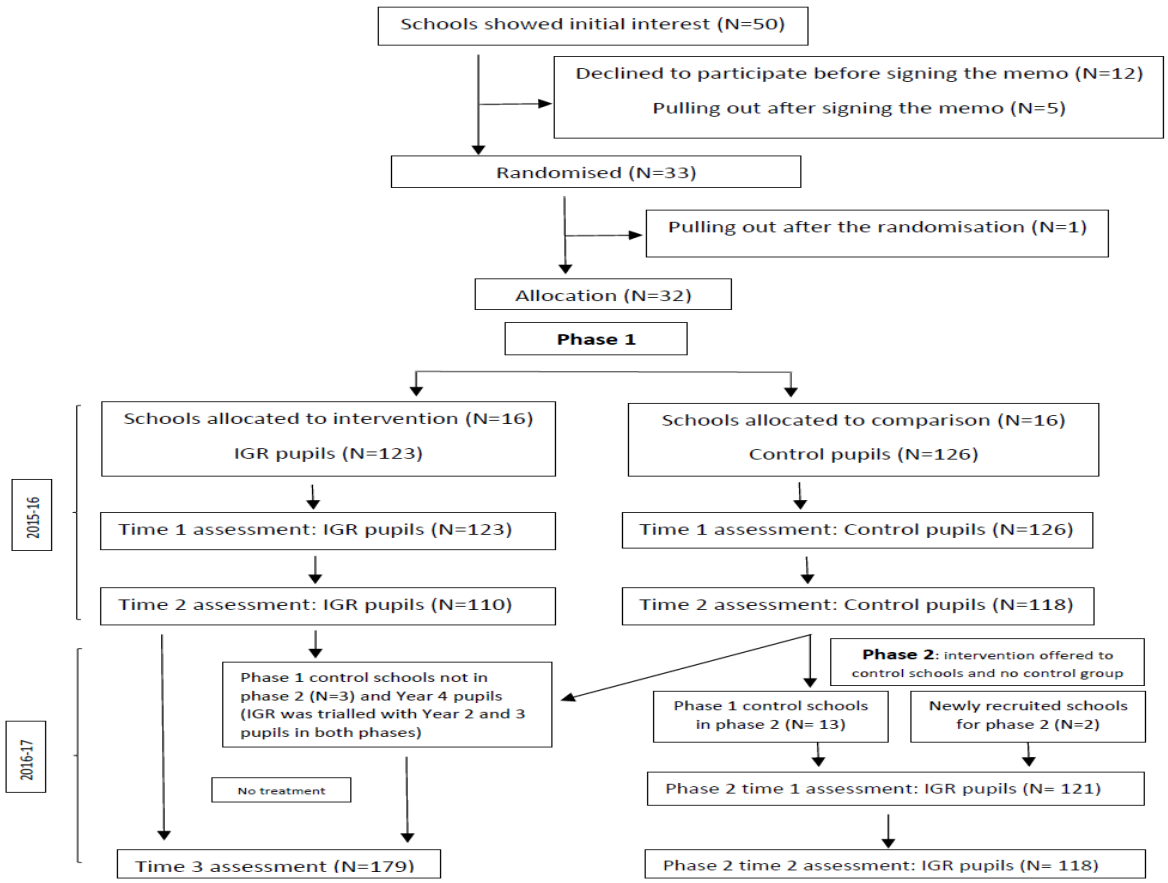
In phase 1, 32 schools participated across 4 Local Educational Authorities in the South West (1), West Midlands (2) and Greater London (1). Of these 32 schools, 16 were intervention schools

involving from 1 to 4 teachers/ classrooms. Altogether there were 31 intervention classes (and 33 control), with each of them having a group identified as in need of wave 2 teaching (4 pupils or in some cases 3). In these groups at time 1 (September 2015) there were 123 IGR intervention pupils and 126 comparison pupils in Years 2 and 3 (whose teachers were asked to continue with typical teaching). At time 2 (July 2016) a small number of pupils were not available to be re-assessed (in most cases they had left the school - numbers in flow chart, Figure 2).

At the end of phase 1 (and during phase 2 – 2016/17), phase 1 intervention schools could decide to continue using the IGR programme or not. In most cases, teachers reported that they were using the programme materials and aspects of the routine, but in a looser way (e.g. delivering fewer sessions and/or not using the full programme routine). All phase 1 intervention pupils were assessed again at time 4 (May 2017).

In phase 2, the intervention was offered to the phase 1 control schools (16 schools) – in phase 1, these schools had no access to the programme. Of these schools, 3 decided not to use the programme due to organisational and staffing difficulties, and then 2 new schools were recruited from another South West local authority area. In sum, in phase 2, there were 15 intervention schools (13 control schools + 2 newly recruited), and there was no control group; there were 33 intervention classes, with each of them having a group identified as in need of wave 2 teaching. At time 1 (September 2016) there were 121 IGR intervention pupils, and 118 were assessed again at time 2 (May 2017) (details in flow chart, Figure 2).

Figure 2. IGR participant flow chart for both phases 1 and 2



Randomisation

Randomisation was applied at school level and took place before the time 1 assessments (September 2015). 33 schools were randomised into the treatment and control groups, with one school deciding to pull out just after the procedure.

Timings of outcome measures

There were 4 assessment times: September 2015 (phase 1, time 1), July 2016 (phase 1, time 2), September 2016 (phase 2, time 3), and May 2017 (phase 1 & phase 2, time 4).

Assessments included individual and whole class assessments (table 2), as below:

Individual assessments

The individual assessments were conducted by visiting Research Associates (RAs) blind to the allocation of schools to intervention and comparison conditions (for the full duration of the project). RAs were recruited from universities close to the participating LEAs and had varied experience of standardised assessments but all had some educational research experience. They received training and support to undertake the assessments. This involved trialling the measures and scoring the tests for which they received feedback from the lead researcher. Individual assessments were conducted with all the identified pupils (IGR and comparison pupils) (see Figure 2 for exact numbers). The RAs administered the following measures:

- York Assessment of Reading for Comprehension (YARC);
- Single Word Reading Test (SWRT);
- 'How I Feel about Reading' (HIFAR); and
- 'How I Feel about My School' (HIFAMS).

York Assessment of Reading for Comprehension (YARC): The YARC test for primary age pupils (GL Assessment Co.) gives separate reading scores for reading accuracy, rate and comprehension for text passages, of which we were interested in accuracy and comprehension. The YARC was selected as it has very satisfactory psychometric characteristics (Snowling et al, 2009). We considered using the Neale Analysis of Reading Assessment (Neale, 1997) because it had simpler administrative characteristics, but decided against it because of it being dated.

For a YARC score to be calculated, a child needed to read two reading passages that represent different levels of difficulty. This meant that when only one passage is read no reading score can be calculated. As the pupils in the IGR and comparison groups were identified by their teachers as delayed readers who would benefit from additional literacy support, in phase 1 only 48% of pupils could read two passages and thereby have a YARC score. This was particularly evident at time 1 (September 2015), and with Year 2 pupils. Analysis of missing values in the data analysis showed that 56% and 44% of YARC scores were missing for the control and intervention groups respectively (this difference is difficult to explain). Though the YARC instructions suggest that in such cases the Early Years YARC test be used, this is a completely different test (with a focus on phonemic awareness) and it was therefore not considered a viable option. It was also found that there were wide discrepancies between Local Authorities in Year 3 IGR children's initial standardised scores.

Single Word Reading Test (SWRT): The SWRT developed by Foster (2007) is part of the YARC test (a tool to select an appropriate starting passage). We used it because we expected that some pupils might not reach the floor of the YARC test and we needed to have a test that could be used by all pupils. The SWRT test has 60 words that a child could either read correctly (word read at sight or analysed), or incorrectly (word omitted or read incorrectly). Correctly read words were awarded one score point. All pupils could access the SWRT. In addition to the SWRT scores for the individually assessed pupils, we used the results of the Hodder Group Reading Test (HGRT) presented below as part of the whole class assessments.

'*How I Feel about Reading' (HIFAR):* The HIFAR covers *reading attitude* (10 items, e.g. 'Do you like word games in class?') and *reading competence* 10 items, e.g. 'Can you work out what a story means') using a 5-point scale, including also two practice items. The scales have satisfactory psychometric characteristics (Chapman & Tunmer, 1995). RAs read the items to the pupils and explained in a non-leading way any confusing items (especially with Year 2 pupils). At time 1 (September 2015), many Year 2 pupils seemed to find HIFAR difficult to follow.

'How I Feel about My School' (HIFAMS): HIFAMS covers the area of school well-being and experiences (7 items, e.g. 'When I think about school I feel…' using in a 3-point scale) using a children-friendly design (illustrations with facial expressions). The scales have satisfactory psychometric characteristics (Allen et al, 2017).

Whole class assessments

Whole class assessments used the Hodder Group Reading Test (HGRT). A whole class reading assessment was used to explore any effects of the delivery of IGR on other class pupils. The HGRT was used for all the pupils in the classes where there was an IGR or comparison group of pupils. This included the pupils who were individually assessed (compared to the YARC, the Hodder test proved to be a more widely accessible test). The HGRT has two versions, of which both version 1 (5:0-9:0 years) and version 2 (7:00-12:00 years) were used. For Year 2 and 3 pupils (6:00-8:00 years old) the tests overlap, and so teachers decided which version they thought each child should take based on sample tests. There are alternative forms for each test of the same difficulty. Form A was used at time 1 and Form B at time 2 (this also applies to the YARC test). The HGRT test version 1 has 40 items and the HGRT version 2 53 items that examine a combination of reading accuracy and comprehension. Version 2 is also more advanced and timed (roughly 35 minutes). To compare scores on HGRT across versions 1 and 2, standardised scores had to be used, not raw scores. All IGR pupils could access one of the HGRT versions. The HGRT has satisfactory psychometric characteristics (Hodder Education, 2000) and even though the norms were dated, it was the best test that could be identified.

The Hodder tests for the whole class were administered by the class teachers. The paper tests were sent to the schools by the research team, accompanied by very detailed step-by-step instructions (and an envelope with prepaid postage). The completed tests were then returned to the research team for scoring.

Table 2. An overview of the programme's assessments

Reading assessments	Description
Individual assessments	
Single Word Reading Test (by visiting RA)	Simple test that captures reading accuracy only and indicates whether YARC test can be used
YARC test (by visiting RA)	Standardised test for reading accuracy/ comprehension
Reading self-concept questionnaire (by visiting RA)	Questionnaire about pupils' reading self-concept
School well-being questionnaire (by visiting RA)	Questions about their school experiences generally
Whole class assessments	
Hodder test (by the class teacher)	Standardised group reading test for the whole class where there is an IGR (treatment/ control) group of pupils

The power analysis and overview of statistical methods is presented in Appendices 7 and 8.

Ethics are discussed in Appendix 9.

5. Key findings from the evaluation

5.1. Experimental findings

Experimental findings are presented in summary here and in detail (including tables) in Appendix 10.

Participants

There were no significant differences in the number taking part in the treatment and control phases, apart from the SEN School Support group: more particularly, there were more children identified for SEN school support in the control schools compared to the treatment schools in phase 1; this might suggest that in phase 1 some treatment schools could perceive the IGR programme as a substitute to SEN School Support provision and so not identify pupils having IGR as eligible for SEN Support.

Main findings

Both in the treatment and control groups, pupils showed progress on the standardised reading test scores between the pre-treatment and post-treatment measurements. But, there was no statistical evidence to conclude that IGR had a positive or negative effect on IGR pupils' outcomes compared to the alternative programmes used in the control classrooms. Effect sizes are mostly close to zero and are never larger than 0.25.

We also explored the interaction between IGR and a number of socio-demographic characteristics of pupils, in order to test the hypothesis that an IGR effect might be different for boys and girls, pupils in Years 2 and 3, native and non-native English speakers, with and without special education needs or Pupil Premium status. Overall, we did not find consistent statistically significant interaction effects between IGR and any of the socio-demographic variables. In phase 1, IGR girls did significantly better on the Hodder test scores than IGR boys, but this effect was not replicated in phase 2 nor with the SWRT test in either phase. Similarly, in phase 1 IGR children with English as an additional language showed better progress compared to native speakers, but again the effect was not replicated in phase 2 or for SWRT scores. Also, in phase 2 (but not phase 1), IGR pupils with the Pupil Premium status did significantly better on the SWRT test (but not on HGRT).

Given a large number of hypotheses we are testing for the interaction effects, some of these effects are likely to be statistically significant just by chance. None of the effects is consistent for both measures of the reading progress and in both phases of the study.

Longer-term effects

We also explored the long-term IGR effects for phase 1 children – defined as 1.5 years after the start of the IGR intervention (phase 1 implementation) and around 9 months after the evaluation ended (IGR might have been used during this period in some form or extent or not). We refer to this assessment as 'time 4'.

The only statistically significant long-term IGR effect is for HGRT. The progress on HGRT between times 1 and 4 was larger in the control schools compared to the treatment schools, and the IGR effect is negative (the effect size is approximately -0.3) Yet, as at time 4 the mean HGRT scores in the control and treatment groups were very similar, so the negative IGR effect is likely due to the lower Hodder scores in the control group at time 1 – suggesting a measurement error for HGRT at time 1.

Rest of the class

We also explored whether IGR affected the outcomes of pupils who were not directly involved in the IGR programme (the rest of the class). These children's progress could be affected by IGR because of changes in the class organisation, the teaching and their teacher's self-efficacy etc. We found then that in both phases IGR showed a small positive effect on the reading progress (as measured by the HGRT) of the non-IGR children. The standardised effect size was 0.13 in Phase 1 and 0.23 in Phase 2. In phase 2 the effect was statistically significant at the 95% significance level ($p = 0.03$).

However, we should be careful with not over-interpreting these results. When we fitted the interaction effect between IGR and gender in the models for the non-IGR pupils, the positive IGR effect was only present for girls. `So, the explanation for this is significantly higher pre-treatment baseline mean HGRT scores for girls in the control group.

In addition, in all conditions (with the exception of girls in the control group) pupils showed approximately the same progress in the HGRT between pre- and post-treatment measures (between 3 and 5 points). Only girls in the control group did not show any progress. A possible explanation is that a measurement error in the control group at time 1 introduced a possible upward bias.

5.2. Process evaluation

Process evaluation methods and findings are presented here in summary and in detail in Appendices 11 and 12.

Methods

14 schools (8 in phase 1 and 6 in phase 2) (mixed range of rural, sub/urban schools), each acting as different cases, were selected across the 4 LEAs based on a combination of differing levels of a school characteristic (percentage of pupils receiving Free School Meals) and a teacher characteristic (different levels of teacher self-efficacy for teaching reading) – the latter was based on a reading teaching self-efficacy questionnaire completed by teachers as part of the training day.

The process evaluation was designed using a realist evaluation framework as discussed above (Pawson & Tilley, 1997). An IGR programme theory was constructed in terms of over 100 context, mechanisms and outcome areas expressed in both positive and negative terms to inform the data collection (questionnaire, interview schedule) and analysis (thematic structure).

In each school one (or more) teacher-led IGR session/s was observed and one (or more) teacher/s was interviewed and completed an IGR context, mechanism and outcomes (CMO) questionnaire based on a shorter version of the programme theory. The questionnaire asked for the perceived interim outcomes of the programme (10 items) and the mechanism (20 items) and context factors (10 items) that had supported them. After each lesson observation, pupils in the intervention groups (IGR groups) were asked about their experiences of the programme.

These data were complemented by 10 online logs completed every fortnight by all the IGR teachers to monitor how they implemented the programme (implementation fidelity). The log asked teachers to summarise their classroom organisation, the number of teacher and TA sessions, the programme's teaching routine, and comment on pupil attainment and attitude. It was reviewed and revised several times during the year to better capture departures from the suggested organisation and methodology. In addition to the log, there were observations from the programme team support visits to schools (conducted at various intervals during the year) and focus group interviews from local review meetings with the teachers at the end of two terms of the IGR programme.

Control teaching was not explored directly through school visits to avoid putting additional pressure on control schools who were on a waiting list to use the intervention. An online survey was sent in November 2015 to the teachers who were in the control group and were continuing with their usual teaching. The survey asked for information about the teaching of reading and the classroom organisation. A second survey was sent at the end of phase 1 (autumn 2016), focusing more on the 4 identified comparison pupils and the use of additional literacy programmes. Two control teachers participated in a follow-up phone interview to discuss their responses in more depth.

School visit interviews were analysed using the constant comparative method at a case level and then again across cases. Review meeting interviews were analysed in the same way but separately from the individual cases. The same thematic structure informed by the programme theory was used [note 2]. Information from the log and the programme team observations was organised in themes by teacher. A summary for each school was initially produced (to synthesise school visit data, and data from the log and the programme team observations), and then a cross case summary was prepared to summarise common themes. Control teaching data were analysed separately and a summary was produced.

Main findings

IGR was implemented with varied fidelity across different schools and teachers. Most common variations observed or reported in the monitoring log included delivering IGR out of the classroom, confusing the teacher and teaching assistant roles, delivering fewer than the suggested sessions, and not following the IGR lesson routine.

IGR proved to be a demanding approach as far as teacher skills were concerned, since it adopts a multi-perspective approach that can be seen as different from the current approach to early literacy that emphasises synthetic phonics and inference. Teachers had mixed views on this: some younger teachers who had been trained with a focus on phonics tended to alter the delivery of IGR slightly to be closer to a more phonics-driven instruction. In a similar way, the story-telling element of IGR for some teachers tended to be altered into a more inference-driven approach to text with teacher questions and pupil responses, in a teaching style closer to the Guided Reading approach that aims to make pupils independent readers. On the other hand, many teachers appreciated the simplicity of IGR that combined a variety of light touch approaches to re-engage pupils in reading.

The IGR organisation was described as 'marginally more demanding' by one teacher, and this seemed to reflect the overall attitude of all interviewed teachers. Most teachers could see the value of keeping all the pupils in the classroom during the intervention, but there were a few teachers who saw a tension between the inclusive aspect of IGR (keeping all pupils in the class) and the difficulty of maintaining concentration in a busy and lively class. Giving the main role to the teacher means also that a teaching assistant has to be available to work with the rest of the class, with some schools reporting issues with teaching assistant availability.

The main issue schools and teachers had was with the number of teacher sessions when there were more than four reading groups in the class. With four reading groups already, this meant that teachers had to fit these four sessions into three days for the period of the intervention. Teachers came up with a variety of solutions with the most common being the delivery of one of the two teacher-led IGR sessions in the classroom but during a school assembly. Many schools made clear that all reading groups and pupils should have an equal entitlement to the teacher's time.

The control teaching data were collected from the phase 1 control schools only (there was no control group in phase 2), using two online surveys sent in autumn 2015 and again in autumn 2016. When control teachers were asked how much time they and their teaching assistants spent with the identified pupils, they reported giving considerable additional time to the identified pupils (figure 2). In addition, a number of literacy programmes was used in control schools, including ReadWrite Inc, Toe by Toe and The Five-Minute Box for Literacy.

5.3 Case studies

The decision to undertake case studies was taken after the trial data collection to examine the relationship between IGR teaching quality and IGR group reading outcomes. The assumption was that IGR targeted teaching had the potential to result in greater reading gains than were found in the trial overall, and that where this did not happen it was due to less than adequate IGR teaching and other factors in the class organisation and school context. This meant that the IGR teaching cases were selected by identifying teachers with high fidelity levels and high pupil mean reading gains, on one hand, and some with low fidelity scores and low or negative gains, on the other. However, it was also necessary to analyse cases where there was higher IGR fidelity but lower reading gains, on the one hand, and low to medium fidelity but higher reading gains on the other. (For more details about design see appendix 13).

Methods

The selection of cases was based on a combination of fidelity and mean group reading gain scores from the This procedure was carried out for phase 2 only, as there was less accurate phase 1 programme observation data. Seven teachers were selected for case analysis.

Main findings

The case analysis of teachers (T) can be summarised as follows:

T1 and T2: high gains can be attributed to the quality of IGR teaching and other supportive factors.

T3: low/no gains can be attributed to specific aspects of IGR teaching – this case also indicates that the fidelity index might not give due weight to the quality of collaborative reading activity.

T4: the low gains can be attributed to the low level of IGR teaching and organisational issues.

T5: the good gains can partly be attributed to other programmes operating during IGR, and partly to IGR teaching, as the fidelity index could not capture appropriate changes in pace. The IGR pupils had also higher initial reading scores compared to other groups in the study.

T6: low gains can partly be attributed to implementation, class organisation and TA-related issues and partly to not using the full sequence and number of teacher/TA-led sessions.

T7: low gains were not only to do with medium IGR fidelity (not high level) but large class, job-share and inconsistent IGR approach between job-sharing teachers.

6. Summary of key points

Experimental evaluation

Participating children in schools using IGR in both phase 1 or 2 made the same degree of progress in reading accuracy/ comprehension, compared to similarly struggling children in control schools (mainly using phonics approaches). The mean reading progress in intervention and control groups was equivalent to 11 months over the 7 months in phase 1, often seen as 'modest impact'; and mean progress in both groups of 14 months over the 7 months of phase 2, often seen as 'useful impact'. There were also no statistically significant changes for reading and school attitude in either the treatment or control group. This suggests that our initial hypothesis that IGR would improve reading gains and attitudes for the IGR group compared to the control group was not supported by the findings.

There were no *consistent* statistically significant interactions between the IGR programme and gender, Year Group, and having English as an additional language (EAL). Pupils having EAL and being identified for Pupil Premium made significantly greater gains with IGR, but these findings were not replicated across phases, measures or levels of significance.

In Phase 1 there was no statistically significant difference in gains between treatment and control classes for non-IGR children. This confirms our initial hypothesis that IGR in the classroom would not have any negative effect on the classroom pupils not having the intervention. In Phase 2, non-IGR children showed somewhat better progress on the Hodder standardised scale in the treatment classes compared to the control classes (d = 0.2). This effect was statistically significant, but this is interpreted as probably due to the high baseline scores for the girls in the control group and possibly because of a measurement error.

For teachers using IGR, their self-efficacy in teaching reading through a self-report measure improved significantly in both phases. Control teachers did not complete this measure, so this change is hard to interpret.

Process evaluation

Overall, participants were enthusiastic about the intervention, the project materials, and accompanying support. Teacher-reported outcomes for IGR pupils included increased confidence, motivation and interest in reading, and improved reading, oral language and social skills. Some teachers were concerned that these gains had not yet transferred outside of the IGR group setting. Most pupils were not worried about being seen in a low attainment group, and did not see IGR as an intervention, but as an exciting classroom activity. Other class pupils were often very interested in the IGR resources, especially the games.

IGR was used with varied fidelity, and many teachers had limited understanding of the theory underpinning the programme, which could partly indicate a training limitation. In phase 1, this

resulted in the programme support team having to produce a table with acceptable and unacceptable variations to programme implementation to advise teachers accordingly. This reflected individual variation in the way the programme was used, and that fidelity was operating along a continuum. Some departures from the suggested methodology were seen in some cases to be justifiable (such as, slowing down the pace of the programme in response to pupils' needs), whereas others were less acceptable (for instance, delivering all programme sessions in withdrawal sessions).

In addition, control schools did not just continue with typical teaching; teachers recognised that control pupils had significant additional needs, so they also had a considerable additional, mainly phonics-based teaching input, making what was being compared to the experimental evaluation varied and complex.

Case studies

Two teaching cases showed that when high reading gains followed high IGR teaching fidelity, these supportive factors were identified, e.g. teacher and pupil enthusiasm, school leader and advisor involvement, teacher understanding the theory and rationale of IGR and the IGR model fitting the pre-existing reading organisational arrangements. When low gains were followed by low fidelity in two other cases, the above factors were not identified.

In cases where low reading gains followed medium to high IGR fidelity, there was evidence of factors that were barriers to reading progress, such as, a mechanical teaching approach that did not engage pupils, having a TA who could not manage the other groups during IGR teaching and unsatisfactory teacher job-sharing arrangements. In the case where quite high reading gains followed low IGR fidelity, there was evidence that the measure of fidelity was affected by changes to teaching which did not affect the otherwise high quality of IGR teaching.

Overall findings

The experimental evaluation indicates that the multi-perspective IGR approach that supports enjoyment of reading resulted in as much reading gain as the more phonics-oriented programmes used in control classes. The process evaluation and case studies illustrate further benefits and some challenges not found in the measured outcomes. This means that IGR might be considered by schools and teachers as in place of the current pattern of targeted interventions that involve more phonics-based programmes delivered by teaching assistants.

The lack of negative effects on reading in non-IGR pupils is a noteworthy finding suggesting that IGR in the classroom could be considered as an alternative to the usual model of offering additional support in withdrawal sessions often led by teaching assistants.

The gains in teachers' perceived self-efficacy in teaching reading following the use of IGR point to possible confidence benefits in teaching reading.

While the process evaluation showed IGR strengths (e.g. teacher enthusiasm and enhanced pupil confidence) and limitations (e.g. insufficient understanding of the theory of IGR teaching), on one hand, the teaching case studies illustrated how useful reading gains depended on IGR teaching fidelity and other supportive factors.

The programme does not have considerable implementation costs apart from the one-off cost of materials; however, it has particular staffing demands (teachers and teaching assistants), and it can be used more effectively in classrooms that have a regular teaching assistant.

Significance of the key findings

IGR adopts a multi-perspective approach that includes phonics, but also enables pupils to engage more deeply with text and allows for comprehension to emerge naturally. Following the Rose (2006) report, schools in England largely use phonics approaches to teach early reading, with explicit phonics teaching showing good results (Wyse & Goswami, 2008).

The experimental evaluation indicates that the multi-perspective IGR approach that supports enjoyment of reading resulted in as much reading gain as the more phonics oriented additional programmes used in control classes (as evidenced by the data collected from control schools). Since we had decided not to intervene with the teaching decisions in the control classes, the trial was not in fact comparing IGR with no additional support but to a programme influenced by the dominance of the phonics approach and driven by the National Curriculum and the assessment requirements.

The IGR trial also found that IGR organisation made it possible for the teacher to deliver targeted (tier 2) teaching to pupils who are delayed in their reading in the regular class, without this having a negative impact on the rest of the class pupils. We take this to mean that IGR might be considered by schools and teachers as an alternative to the current pattern of targeted interventions that involves more phonics-based programmes delivered in most cases by teaching assistants. These findings are relevant to teachers, advisers and policy makers who are looking for more inclusive approaches for targeting pupils in years 2 and 3 who are delayed in their reading.

Implications about how additional support is organised

The IGR organisation that enables teachers to offer targeted (tier 2) teaching in a 'Quality First' (tier 1) setting proved to be challenging but viable. This has implications about the way additional provision is organised for pupils identified as in need of targeted support. It particularly shows how it is practically possible for the teacher to take responsibility for the learning of all pupils, even by offering extra time to some most needing it, without hindering the learning of the rest of the class.

This model for the organisation of additional support (group organisation, coordinated teacher-TA collaboration and well-prepared materials) could be extended beyond reading to other areas of learning, e.g. aspects of mathematics or science. Future research and development might explore this approach to inclusive targeted support beyond the teaching of reading.

Areas for future research and development

As a first step after completing the study, there is scope to explore how and whether IGR is being used by teachers in participating schools with the restrictions of the RCT protocol removed. It is likely that teachers would mainly use the programme materials in a loose way, but there is still value in exploring the reasons behind their decisions.

Future studies could also be designed with greater focus on teacher professional learning about the principles of IGR and more focussed training and coaching of IGR-related teaching skills.

With regards to programme development, the IGR programme developer, building on teacher feedback, has designed a synthetic phonics game that could readily be added to future versions of IGR (the version trialled here had an analytic phonics component). This game is story-specific and fits with the existing programme materials.

6.1. Implications for teaching practice

The IGR approach to reading

The experimental evaluation indicates that the multi-perspective IGR approach that supports enjoyment of reading resulted in as much reading gain as the more phonics-oriented programmes used in control classes. This means that IGR might be considered by schools and teachers as an alternative to the current pattern of targeted interventions that involve more phonics-based programmes delivered by teaching assistants. This can be related on the one hand to the teacher-led classroom-based delivery of IGR discussed in the next sections, and on the other to the nature of the IGR methodological approach to reading discussed here.

Following the Rose (2006) report, schools in England largely use phonics approaches to teach early reading, with explicit phonics teaching often shown to have good results (Henbest & Apel 2017; Wyse & Goswami 2008). However, as some teachers in this study stressed, one approach cannot be relevant to all pupils, and IGR can be used with those pupils for whom a mainly phonics approach has not resulted in enough progress. IGR can thus be used alternatively to single phonics approaches, and especially where phonics has not helped struggling pupils re-engage with reading and regain their confidence as readers.

Lovett et al (2017) discuss the importance of multi-perspective remedial programmes that have the potential to address a number of issues beyond phonological difficulties. The implication of this is that research-informed teachers would select the method they feel better suits their teaching style and pupils' needs, whether this method is phonics-centric or not. Our findings suggest that some teachers feel comfortable in using single phonics approaches, whereas others are attracted to multi-perspective approaches (such as IGR). This is particularly evident in the way storytelling was used by the teachers in the study, with some founding it very challenging, and others experiencing it as a natural activity. The implication is that the selection of an appropriate method for the teaching of reading should not be a simple matter of policy, but be research-informed and involve teacher decision-making.

IGR and EAL pupils

Pupils with English as an additional language (EAL) made higher reading progress in phase 1 treatment schools compared to control schools (though not at the $p<0.05$ level of significance), but the same reading progress in phase 2. As this finding was not consistent across phases, further research is needed to explore the potential effects of IGR for this particular group. There are indications though that the IGR approach which is based on books with colourful illustrations, storytelling and discussion, and on games that scaffold oral language, vocabulary and social skills might be particularly suitable for EAL pupils. In addition, the simplicity and fun element of IGR can reduce the pressure of learning that is likely to be experienced by an EAL pupil. This was noted by teachers in schools with a high proportion of EAL pupils in their IGR groups. A few teachers also used IGR materials with non-monitored EAL pupils (who in some cases joined the school later in the year), acknowledging the potential value of the programme for these pupils. Further research is needed to determine the extent to which IGR can be relevant to this particular pupil group. The same applies to a similar finding about a positive interaction for pupils identified for Pupil Premium in phase 2.

IGR organisation in relation to the way additional support is often organised

With reference to those classroom pupils not having IGR teaching, the change in reading outcomes (measured by the HGRT) was approximately the same in the control and treatment groups across phases. This is consistent with IGR teaching of a sub-group not affecting how other pupils progressed either positively or negatively. Similar findings have been reported in a recent EEF report (Patel et al., 2017) but the intervention was delivered in TA-led pull-out sessions. This is a notable finding suggesting that there is an alternative to the usual model of offering additional support.

A literature review was conducted by the authors to explore the nature of additional support for children struggling to learn to read. It was found that the people responsible for teaching targeted tier 2 interventions that were evaluated in the UK were mostly TAs and the sessions were often delivered out of the regular class (e.g. See et al, 2015; Clarke et al, 2010; Duff et al, 2008; Hatcher et al, 2006). This can be indicative of how additional support is often organised in UK primary schools. Although there is a risk of over-generalising to situations where appropriate support is provided by TAs, the use of pull-out sessions and a reliance on people other than the classroom teacher to offer additional support can invite 'a separation effect' (EEF, 2015, p. 15), with certain pupils spending considerably less time with the teacher and having fewer opportunities for peer interaction. This raises the question of whether Quality First programmes are sufficiently differentiated for those struggling to learn, and why it is that programmes at tier 2 are offered as supplementary programmes.

The implication of this evaluation is that it is practically possible (and does not bring any negative effects) for the teacher to offer tailored targeted wave 2 support in the Quality First teaching setting, during a whole-class teaching session. This would allow wave 2 pupils to access tailored teaching and spend quality time with their teacher and peers. As this is a matter pertinent to broader issues about the relationship between what is offered to all (general provision) and what to some (additional provision), the IGR organisational model could also be applied outside the area of reading to any other topic that can be taught in a group-based whole-class organisation.

Teacher self-efficacy

As far as the teachers are concerned, the change in self-efficacy for the treatment teachers between the training day (October 2015) and the review meetings (July 2016) was found to be statistically significant (control teachers did not complete a self-efficacy questionnaire). The same applied to teachers in phase 2.

This suggests that the IGR programme can be associated with teachers becoming more confident in their literacy teaching. The significance of this lies in that IGR organisation made it possible for teachers to work with their pupils who struggled the most without having to leave the classroom; often this group is taught by TAs out of the classroom. Although teaching these children can be particularly challenging, teachers reported in their interviews and CMO questionnaire that they felt they took responsibility for the learning of every pupil in their class and that they could monitor pupil progress in greater detail. In addition, IGR gave better structure to the way provision was organised for all (in the form of Guided Reading or similar approach) and involved a number of strategies that could be used with all pupils – for instance a few teachers used the Lotto game (aiming to make pupils familiar with unknown vocabulary before reading) in whole class teaching.

These factors, alongside the training and support from reading specialists and Literacy Advisers, could have contributed to making teachers more confident in their literacy teaching.

Attitudes to the programme

Reading and school attitudes as measured by the HIFAR and HIFAMS questionnaires remained largely unchanged across both phases. Interview data from the process evaluation suggests that IGR pupils were enthusiastic about IGR materials and were engaged in their reading, so perhaps a questionnaire focusing on the enjoyment of reading would be more relevant than the scales used in the study to capture these changes. Assessing these affective aspects of reading and schooling is challenging (for instance McKenna et al., 1995), and perhaps some alternative form of assessment might have been more sensitive to the changes that arose during the period of the study.

There was also evidence that pupils, with few notable exceptions, were not concerned about being visible in the IGR group, a low attainment group, and that being in the IGR group was often seen as a privilege because of the unusual activities, materials and games. This indicates that there is no evidence that, for pupils of this age in this kind of classroom context, temporary reading ability grouping leads to the devaluation of the identified pupils. Assumptions about grouping, as practised in IGR, as leading to devaluation and stigma underpin the general rejection of ability grouping advocated in some inclusive pedagogy perspectives (Florian & Black-Hawkins, 2011). This evaluation illustrated that the inclusive features of IGR teaching are compatible with the temporary reading ability grouping of some pupils. It is indicative that other pupils were often envious of the programme's activities, especially the games.

What was compared in the trial: IGR fidelity and control teaching

As discussed, on the one hand IGR was implemented with varied fidelity, and on the other hand, control schools offered intensive additional support to control pupils. Thus, what was actually being compared in the trial was difficult to pinpoint. This issue can be understood in terms of the concept of 'complex intervention'. Based on realist principles, Moore et al. (2015) define complex interventions as interventions that comprise:

> '…multiple interacting components, although additional dimensions of complexity include the difficulty of their implementation and the number of organisational levels they target' (p. 1).

IGR can be seen as a complex intervention, involving a multi-perspective teaching methodology with a particular teaching routine, and strictly separate but related roles for teachers and teaching assistants. IGR was also implemented in the regular class and not in pull-out sessions by people other than the classroom teacher. In this sense, the programme was a real-world experiment where control over implementation fidelity and the study's protocol had to be balanced with the constantly changing situation in classrooms and schools.

Discussing the matter of fidelity to a complex intervention, Moore et al. (2015) note that fidelity is best seen as a matter of degree rather than as a fixed quality. In the study, most teachers tried to stay faithful to the study's protocol, but some found this particularly difficult. A striking example was some teachers' difficulty in keeping their pupil grouping unchanged for the duration of the year (evident in phase 1), since the usual practice is to change the composition of groups according to pupil progress (with ability-based groups) or in response to other issues such as personality clashes. In terms of teacher fidelity to the IGR teaching, it was clear that the requirements of the National Curriculum were having an effect on teachers: two phase 1 teachers, for instance, added activities

to the IGR routine in accordance with the assessment requirements for phonics and comprehension. This is also illustrated by the need to produce a table with acceptable and not acceptable variations to IGR to advise the teachers accordingly. The table suggested that fidelity could operate differently in terms of structure and adherence to the study's protocols. This is consistent with findings from other studies, for example Gorard et al. (2015) and See et al. (2015), where the fidelity of school-based intervention implementation was found to have varied. Most of these issues were less (but still) evident in phase 2.

Also, young or inexperienced teachers (trained mainly with a focus on synthetic phonics) found it particularly difficult to use the programme with more depth of understanding. While many teachers spent a lot of time on preparing for the practical, organisational aspects of the programme, fewer teachers devoted time to reflect on the actual teaching approach of IGR and the strategies involved (which is also relevant to their busy timetables). This is evident from the teacher interviews where few teachers could discuss the IGR methodology and approach to learning in some depth.

In addition, we examined the teaching in the comparison classes in some detail, and we found that typical teaching was not just about usual class teaching but had included intensive additional support of various forms (teacher and TA-led, in and out of the classroom) and often the use of other literacy programmes – in many cases with a focus on synthetic phonics (such as Toe by Toe). This is a matter evident in other studies as well; Vaughn et al (2016) for example when reporting the evaluation of a USA school-based reading intervention for primary pupils, noted:

> 'After students were identified with significant reading comprehension problems and were randomized to treatment and comparison conditions, the schools decided to provide their own interventions to students in the comparison condition. Because students were so far behind, it was unethical to ask them not to provide the intervention' (p. 40).

Based on our findings about the control teaching, we can conclude that this applied in this study too. As we had decided not to intervene with the teaching decisions in the control schools (largely influenced by the dominance in England of the phonics approach), we were in fact comparing IGR not just to usual teaching arrangements but to an intensive programme of (mainly phonics-oriented) support, driven by the National Curriculum and the assessment requirements. However, we do not interpret this as the teachers consciously compensating for these pupils not receiving the IGR approach (being in the control group) by offering unusual levels of additional support, i.e. a *compensation effect*, as illustrated by Patel et al. (2017) in a recent EEF literacy programme trial.

This decision not to intervene might have led to less control over what IGR was actually being compared with. Some studies, indeed, have tried to avoid this occurring. For example:

> 'Remedial instruction in a one-to-one setting by a reading specialist was discouraged, both during and outside school hours' (Oostdam et al, 2015, p. 435).

However, we believed that this was not the ethical thing to do. Our decision also allowed us to compare IGR to teaching as it really occurs in schools that are seeking solutions and not to a situation artificially created for research purposes; we hoped this would give us a deeper insight into how the programme works.

Another way of seeing this is in terms of the distinction between *efficacy* and *effectiveness* trials. Efficacy is about whether the intervention can work under ideal conditions, while effectiveness is

about working when used in the real world. This distinction links with that between internal validity of a trial and its external validity, i.e. whether the intervention can be generalised to real world conditions. Some authors see a trade-off between these two types of validity (Streiner, 2002), such as the balance between exerting greater experimental control and the relevance of the intervention to real teaching conditions. In terms of Streiner's continuum between efficacy and effectiveness trials, the IGR evaluation was clearly more towards the effectiveness end of the continuum.

6.2. Limitations

Identification

The study's identification procedures adapted from Speece (2011) were adopted as a quick, simple, cost-effective and research-evidenced approach that did not require teachers to administer a standardised assessment or even have to refer to one. However, the lack of reference to standardised assessment scores also had the consequence that a few IGR groups included children with standardised reading scores already above the mean – or pupils with very varied abilities. This was related to the fact that some classrooms had high mean reading levels and few struggling pupils (perhaps less than 4), whereas others had low mean reading levels and many struggling pupils (more than 4). The latter case did not pose a problem to the study as IGR is specifically designed for struggling pupils; in such cases, teachers selected 4 pupils to be individually monitored for the evaluation, and they were encouraged to use IGR with other non-individually monitored pupils as they saw fit. Many teachers followed this advice. However, in the former case of the high achieving classrooms identification for the programme was in some cases seen as problematic for effectively teaching the group and choosing appropriate materials for all. A variety of solutions was sought and the programme team had to step in to assist the teachers.

This matter was more evident in some Local Authorities than in others, reflecting differences in reading attainment across England. To avoid similar issues, future studies should consider introducing a cap of around 85 in standardised score terms.

Issues with the YARC, HGRT and attitude measures

The YARC test was chosen as the main reading test of the study. However, only a small number of pupils could complete the baseline assessments (42% of the control pupils in September 2015). On the one hand, this was due to the test requirement that a reading score requires the completion of two whole reading passages and accompanying questions (as opposed to one for similar tests). Using two passages can improve the accuracy of the test results, but it proved a difficult task for those struggling to read for one reason or another. However, the high level of missing YARC scores could also be attributed to poor RA assessment administration. Decision-making involved with the YARC is complicated and requires previous experience with assessments. As the same assessors were involved in phase 1 and 2, this is more relevant to phase 1. Yet, even in phase 2 where the study's assessors already had experience of two assessment times, 33% of pupils could not complete the test at baseline (September 2016). This suggests that, even with considerable previous experience, using the YARC can result in high percentages of missing values for this particular pupil group.

As a result, despite the resources and time invested in the test, YARC scores could not be used except indicatively in the analysis because of the high proportion of missing values. As the research

team realised early on that not all pupils would be able to have a YARC score, an SWRT score was also used for all IGR pupils (who also took the Hodder test together with their classmates). The analysis then relied largely on the SWRT and Hodder tests.

The Hodder test was delivered by the class teachers, with specific instructions and support. Being teacher-delivered the Hodder test proved to be a simple and cost-effective test. It should be noted that the paper tests were not scored in schools but were posted unscored to the research team. However, it seems that a few teachers had difficulties in following the procedures, and there were some measurement problems (evident at time 1, and especially in the control classrooms) that affected some of the analyses, as discussed in the findings chapter. This could raise some questions about the reliability and consistency of a test that was not independently delivered. There were also practical difficulties with a few schools being slow in administering and dispatching the paper tests.

Finally, the attitude scales measuring pupils' reading and school self-concept used in the study recorded no progress between assessment times. However, teachers reported in their interviews and CMO questionnaire that their IGR pupils became a lot more confident in reading and offered some examples, such as pupils who used to hide their books so they would not have to read at home, but after using IGR they wanted to read the stories to their parents. There was converging evidence on this (from teacher and pupil interviews, CMO questionnaires, and review meetings) across phases. Yet, this reported improvement in pupil attitudes was not picked up by any of the scales used, and one could question their sensitivity and relevance of these measures for this pupil group. Another possible factor could be that more time was needed for changes in pupil attitudes to become evident.

7. Future steps and conclusion

Post-project adaptations of IGR in a variety of schools

As a first step after completing the study, there is scope to explore how and whether IGR is being used after the end of phase 1 and 2 by teachers in schools with the restrictions of the RCT protocol removed. When planning for the time 3 assessments, phase 1 teachers were asked whether they had continued using the programme after the phase 1 implementation, and most reported that they used the materials in a loose way, but not necessarily with the suggested routine and/or organisation. During the review meetings, phase 1 and 2 teachers were asked about their future IGR plans with most reporting that they planned changes to make IGR easier to use in their schools. For example, some teachers were planning to use IGR with groups of 6 rather than 4 pupils, and a few were considering the TAs having a more central role to the programme delivery. Others were more worried about the lack of synthetic phonics, inference and writing tasks that are all requirements of the National Curriculum.

Designing a study about how schools actually make use of the programme after the trial phases, will give insight into the particular challenges associated with IGR, either in relation to the training and support, the organisation model, or the programme's methodological approach to reading. These data could inform programme development and also indicate directions for further research. Data could be collected by an online survey, accompanied by a school visit to observe the actual arrangements in place (with a number of teachers and schools having already expressed their interest). Even if schools are using the programme materials in a looser way or as a Guided Reading resource, there is still value in exploring the reasons behind such decision-making.

Bridging activities between IGR and rest of class

IGR organisation brings tier 2 targeted teaching into the Quality First setting. Yet, there is not much continuity between IGR group activities and the activities of the other groups. The literature that examines the Response to Intervention (RTI) or 3-tier model often suggests that teaching offered in higher tiers (2/3) might not have clear links to the Quality First teaching (e.g. Fien et al, 2015; Jaeger, 2016). This is not surprising as pupils requiring wave 2/3 teaching might not have the same needs as pupils who respond well to Quality First teaching. However, building clear links between Quality First and additional teaching, could on the one hand help teachers better organise their teaching for all pupils, and on the other, help pupils have a better understanding of their progress.

Jaeger (2016) describes an attempt to bring wave 2 closer to Quality First (or wave/tier 1):

> 'After a Tier 1 (classroom-based) unit on character analysis […], fourth graders who struggled on the end-of-unit assessment attended a related Tier 2 class. This class did not simply replicate the Tier 1 curriculum. Rather than focusing immediately on book characters, students drew pictures of, reflected on, discussed, and wrote about their own traits and those of friends and family members. Only then did discussion shift to characters from a picture book read aloud' (p. 186).

In Jaeger's (2016) study, pupils who could not fully access the Quality First teaching (tier 1) session, were withdrawn just after the session and had a second tier 2 session with the same content but simplified and adjusted to their particular needs. This is a way to highlight the continuity and interconnectedness of instruction.

IGR could be used as the centre of such a whole-class model of organising provision, especially since the programme is designed to be delivered by the class teacher during a whole-class session. At least one phase 2 school attempted to use IGR this way, incorporating the programme into their current Group Reading organisation (influenced by principles of reciprocal teaching), and building some links between the other groups and IGR. This was mostly done by slightly adjusting the IGR routine, adding a painting and writing activity that was common across the class. Although in this example the links between the activities of the groups were superficial, one could argue that building explicit links between differentiated group activities has the potential to make additional wave 2 teaching a natural part of whole-class teaching and give structure to provision for all pupils.

Future programme development

Many teachers, and especially those trained in a more phonics way, reported being concerned by the absence of synthetic phonics from the programme. IGR incorporates an analytic phonics element to its methodology based on onset and rhyme. As, the research evidence is currently inconclusive as to the primacy of either phonics approach (Henbest & Apel 2017; Wyse & Goswami 2008), IGR was piloted and trialled without synthetic phonics. However, to address these concerns the programme developer has designed a synthetic phonics game that could readily be added to future versions of IGR. This game is story-specific and in this way fits with the other programme materials.

Other teachers insisted that inference tasks and writing should be incorporated into the programme. However, as the main goal of the programme is to enable struggling readers to learn to read, these activities would not be relevant to the programme and can be practised at other times. IGR has been designed specifically to support struggling pupils who require a relaxed, slow-paced to re-engage with reading and regain their confidence. In addition, as some writing activities are already part of

the TA-led IGR sessions, teachers who are worried about their children's writing could focus on strengthening this aspect of IGR.

Appendix 14 further discusses lessons learned by the programme team from the phase 1 and 2 support school visits.

Final conclusions

Phonics and the teaching of reading

In conclusion, the study's findings show that IGR, an approach to reading that incorporates multiple research-based perspectives (including analytic phonics), trialled with Year 2 and 3 pupils who are struggling to read in primary schools across England, can bring the same results as the currently dominant synthetic phonics approach. This has important policy implications, as the assumptions of the Rose (2006) report, which established synthetic phonics in England, have been questioned (e.g. Wyse & Styles, 2007).

The IGR study suggests that approaches using phonics alongside other perspectives have the potential to bring similar results to a single phonics teaching approach, with the advantage that they can better support the enjoyment of reading as is shown by the IGR process evaluation. The IGR approach could also be used with pupils for whom phonics teaching has failed to bring any results.

The implication of this is that teachers be able to exercise some autonomy in deciding which approach to reading for struggling readers better suits their teaching style and their pupils' needs. Clearly there is a need for further studies using IGR building on the current one. The suggestion is that policy should reflect research conclusions, encompassing other approaches and being open to informed teacher decision-making.

Implications about how additional support is organised

Also, the IGR organisation that enables teachers to offer targeted tier 2 provision in a 'Quality First' setting proved to be challenging but viable. This has implications about the way additional provision is organised for pupils identified as being in need of tier 2 support. It particularly shows how it is practically possible for the teacher to take responsibility for the learning of all pupils – even by offering extra time to some most needing it – without bringing any harm to the rest of the class.

Such an inclusive model could be extended beyond reading to other areas of learning that can be taught using a group-based organisation, e.g. mathematics. Further research work could explore the feasibility of the discussed organisation beyond the teaching of reading.

8. References

Allen, K., Marlow, R., Edwards, V., Parker, C., Rodgers, L., Ukoumunne, O. C., et al. & Ford, T. (2017). 'How I Feel About My School': The construction and validation of a measure of wellbeing at school for primary school children. *Clinical child psychology and psychiatry*, 1-17.

Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., & Wagner, R. K. (2014). To Wait in Tier 1 or Intervene Immediately: A Randomized Experiment Examining First-Grade Response to Intervention in Reading. *Exceptional Children, 81*(1), 11–27.

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., Hulme, C. (2008). Improving early language and literacy skills: differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry*, 49(4), 422–432.

Brooks, G. (2016). *What works for pupils with literacy difficulties*. The Dyslexia-SpLD Trust.

Campbell, M. J., & Walters, S. J. (2014). *How to design, analyse and report cluster randomised trials in medicine and health related research*. Chichester: John Wiley & Sons.

Campbell, M. L., Helf, S., & Cooke, N. L. (2008). Effects of adding multisensory components to a supplemental reading program on the decoding skills of treatment resisters. *Education and Treatment of Children*, 31(3), 267-295.

Chapman, J. W., & Tunmer, W. E. (1995). Development of young children's reading self-concepts: An examination of emerging subcomponents and their relationship with reading achievement. *Journal of Educational Psychology*, 87(1), 154.

Charlton, B. (2005). Educational Games: A Technique to Accelerate the Acquisition of Reading Skills of Children with Learning Disabilities. *International Journal of Special Education*, 20(2), 66–72.

Christensen, C. A., & Bowey, J. A. (2005). The Efficacy of Orthographic Rime, Grapheme–Phoneme Correspondence, and Implicit Phonics Approaches to Teaching Decoding Skills. *Scientific Studies of Reading*, 9(4), 327–349.

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating Children's Reading-Comprehension Difficulties: A Randomized Controlled Trial. *Psychological Science*, 21(8), 1106–1116.

Clay, M. M. (1994). Reading Recovery: The Wider Implications of an Educational Innovation. *Literacy, Teaching and Learning*, 1(1), 121-141.

Comaskey, E. M., Savage, R. S., & Abrami, P. (2009). A randomised efficacy study of Web-based synthetic and analytic programmes among disadvantaged urban Kindergarten children. *Journal of Research in Reading*, 32(1), 92–108.

CPRE (2014). Evaluation of the i3 Scale-Up of Reading Recovery, Year Two Report, 2012-13. *CPRE Research Reports*. Retrieved from http://repository.upenn.edu/cpre_researchreports/2

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(1), 29–46.

DCSF (2008). *Letters and Sounds: principles and practice of high quality phonics, Phase One teaching programme*. London: National Strategies.

DfE (2011). *Evaluation of Every Child a Reader (ECaR), Research report DFE-RR114*. London: DfE.

DfE (2013). *The national curriculum in England: Key stages 1 & 2 framework document*. London: DfE.

DfE (2016). *Phonics screening check and Key stage 1 assessments in England*. London: DfE.

DfE (2017). *SFR49 2017 KS1 National tables*. London: DfE.

Di Stasio, M. R., Savage, R., & Abrami, P. C. (2012). A follow-up study of the ABRACADABRA web-based literacy intervention in Grade 1: ABRACADABRA: a follow-up study. *Journal of Research in Reading*, 35(1), 69-86.

Doyle, M. A. (2013). Marie M. Clay's theoretical perspective: A literacy processing theory. In D. E. Alvermann, N. J. Unrau, & R. B. Ruddell (Eds.), *Theoretical models and processes of reading*. Newark, DE: International Reading Association.

Duff, F. J., Fieldsend, E., Bowyer-Crane, C., Hulme, C., Smith, G., Gibbs, S., & Snowling, M. J. (2008). Reading with vocabulary intervention: evaluation of an instruction for children with poor response to reading intervention. Journal of Research in Reading, 31(3), 319–336.

EEF (2015). *Making Best Use of Teaching Assistants: Guidance Report*. London: EEF.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic Phonics Instruction Helps Students Learn to Read: Evidence from the National Reading Panel's Meta-Analysis. *Review of Educational Research*, 71(3), 393–447.

Elbaum, B., Vaughn, S., Tejero Hughes, M., & Watson Moody, S. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, 92(4), 605–619.

Fien, H., Smith, J. L. M., Baker, S. K., Chaparro, E., Baker, D. L., & Preciado, J. A. (2011). Including English Learners in a Multitiered Approach to Early Reading Instruction and Intervention. *Assessment for Effective Intervention*, 36(3), 143–157.

Fien, H., Smith, J. L. M., Smolkowski, K., Baker, S. K., Nelson, N. J., & Chaparro, E. (2015). An Examination of the Efficacy of a Multitiered Intervention on Early Reading Outcomes for First Grade Students at Risk for Reading Difficulties. *Journal of Learning Disabilities*, 48(6), 602–621.

Florian, L. & Black-Hawkins, K. (2011). Exploring Inclusive Pedagogy. *British Educational Research Journal*, 37 (5), 813-28.

Foster, H. (2007). *Single word reading test 6-16*. London: GL Assessment Limited.

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of Treatment Approaches for Children and Adolescents with Reading Disabilities: A Meta-Analysis of Randomized Controlled Trials. *PLoS ONE*, 9(2), e89900.

Gerde, H. K., & Powell, D. R. (2009). Teacher Education, Book-Reading Practices, and Children's Language Growth across One Year of Head Start. *Early Education and Development*, 20(2), 211–237.

Gorard, S., Siddiqui, N., & See, B. H. (2015). An evaluation of the 'Switch-on Reading' literacy catch-up programme. *British Educational Research Journal*, *41*(4), 596–612.

Griffiths, Y., & Stuart, M. (2013). Reviewing evidence-based practice for pupils with dyslexia and literacy difficulties. *Journal of Research in Reading*, 36(1), 96–116.

Henbest, V. S., & Apel, K. (2017). Effective Word Reading Instruction: What Does the Evidence Tell Us? *Communication Disorders Quarterly*, 1-9.

Hodder Education (2000). *Hodder Group Reading Tests (HGRT) II*. London: Hodder Education.

Hatcher, P. J., Hulme, C., Miles, J. N. V., Carroll, J. M., Hatcher, J., Gibbs, S., Smith, G., Bowyer-Crane, C., & Snowling, M. J. (2005). Efficacy of small group reading intervention for beginning readers with reading-delay: a randomised controlled trial: Efficacy of small group reading intervention. Journal of Child Psychology and Psychiatry, 47(8), 820–827.

Isbell, R., Sobol, J., Lindauer, L., & Lowrance, A. (2004). The Effects of Storytelling and Story Reading on the Oral Language Complexity and Story Comprehension of Young Children. *Early Childhood Education Journal*, 32(3), 157–163.

Jaeger, E. L. (2016). Intensity of Focus, Richness of Content: Crafting Tier 2 Response to Intervention in an Era of the Common Core. *The Reading Teacher*, 70(2), 179–188.

Johnston, R. S., & Watson, J. E. (2004). Accelerating the development of reading, spelling and phonemic awareness skills in initial readers. Reading and Writing, 17(4), 327–357.

Koutsouris, G., & Norwich, B. (2018). Are null RCT findings a secure indicator of programme failure? Exploring the context of implementing the Integrated Group Reading (IGR) programme drawing on teacher case studies. *British Educational Research Journal* (currently under peer review).

Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the Effectiveness of Two Theoretically Motivated Computer-Assisted Reading Interventions in the United Kingdom: GG Rime and GG Phoneme. *Reading Research Quarterly*, 48(1), 61–76.

Lovett, M. W., Frijters, J. C., Wolf, M., Steinbach, K. A., Sevcik, R. A., & Morris, R. D. (2017). Early intervention for children at risk for reading disabilities: The impact of grade at intervention and individual differences on intervention outcomes. Journal of Educational Psychology, 109(7), 889.

McKenna, M. C., Kear, D. J., & Ellsworth, R. A. (1995). Children's attitudes toward reading: A national survey. *Reading research quarterly*, 934-956.

Miller, D., Topping, K., & Thurston, A. (2010). Peer tutoring in reading: The effects of role and organization on two dimensions of self-esteem. *British Journal of Educational Psychology*, 80(3), 417–433.

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D. & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*, 350:h1258.

Muter, V., Hulme, C., Snowling, M.J., and Stevenson, J. (2004). Phonemes, rimes and language skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, 40, 663-681.

NELP (2008). *A Scientific Synthesis of Early Literacy Development and Implications for Intervention*. Jessup: National Early Literacy Panel.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: broader language skills contribute to the development of reading. *Journal of Research in Reading*, 27(4), 342–356.

Neale, M. D. (1997). Neale Analysis of Reading Ability II: Second Revised British Edition. Windsor, England: NFER-Nelson.

NRP (2000). Report of the National Reading Panel: Teaching Children to Read. Washington DC: National Institute of Child Health and Human Development.

Oostdam, R., Blok, H., & Boendermaker, C. (2015). Effects of individualised and small-group guided oral reading interventions on reading skills and reading attitude of poor readers in grades 2–4. *Research Papers in Education*, 30(4), 427–450.

Patel, R., Jabin, N., Bussard, L., Cartagena, J., Haywood, S., & Lumpkin, M. (2017). *Switch-on Effectiveness Trial: Evaluation report and executive summary, May 2017*. London: EEF.

Peck, J. (1989). Using Storytelling to Promote Language and Literacy Development. *The Reading Teacher*, 43(2), 138–141.

Pawson, R. & Tilley, N. (1997). *Realistic Evaluation*. London: Sage.

Raffaele Mendez, L. M., Pelzmann, C. A., & Frank, M. J. (2016). Engaging Struggling Early Readers to Promote Reading Success: A Pilot Study of Reading by Design. *Reading & Writing Quarterly*, 32(3), 273–297.

Streiner, D. L. (2002). The 2 E's of research: efficacy and effectiveness trail. *Canadian Journal of Psychiatry*, 47(5), 553-556.

Rose, J. (2006). *Independent review of the teaching of early reading: Interim report*. London: DFES.

See, B. H., Gorard, S., & Siddiqui, N. (2015). Best practice in conducting RCTs: Lessons learnt from an independent evaluation of the Response-to-Intervention programme. *Studies in Educational Evaluation*, 47, 83–92.

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., et al. (2009). YARC: York Assessment of Reading for Comprehension. London: GL Assessment.

Speece, D. L., Schatschneider, C., Silverman, R., Case, L. P., Cooper, D. H. & Jacobs, D. M. (2011). Identification of reading problems in first grade within a response to instruction framework. *Elementary School Journal*, *111*, 585–607.

Sulzby, E. (1985). Children's Emergent Reading of Favorite Storybooks: A Developmental Study. *Reading Research Quarterly*, 20(4), 458–481.

Topping, K. J. (2014). Paired Reading and Related Methods for Improving Fluency. *International Electronic Journal of Elementary Education; Kutahya*, 7(1), 57–69.

Topping, K. J., & Lindsay, G. A. (1992). Paired reading: a review of the literature. *Research Papers in Education*, 7(3), 199–246.

Topping, K., Miller, D., Thurston, A., McGavock, K., & Conlin, N. (2011). Peer tutoring in reading in Scotland: thinking big: Peer tutoring in reading. *Literacy*, 45(1), 3–9.

Valencia, S. W., & Sulzby, E. (1991). Assessment: Assessment of Emergent Literacy: Storybook Reading. *The Reading Teacher*, *44*(7), 498–500.

Vaughn, S., Solís, M., Miciak, J., Taylor, W. P., & Fletcher, J. M. (2016). Effects from a Randomized Control Trial Comparing Researcher and School-Implemented Treatments with Fourth Graders with Significant Reading Difficulties. Journal of Research on Educational Effectiveness, 9(sup1), 23–44.

Walton, P. D., Walton, L. M., & Felton, K. (2001). Teaching rime analogy or letter recoding reading strategies to prereaders: Effects on prereading skills and word reading. *Journal of Educational Psychology*, 93(1), 160–180.

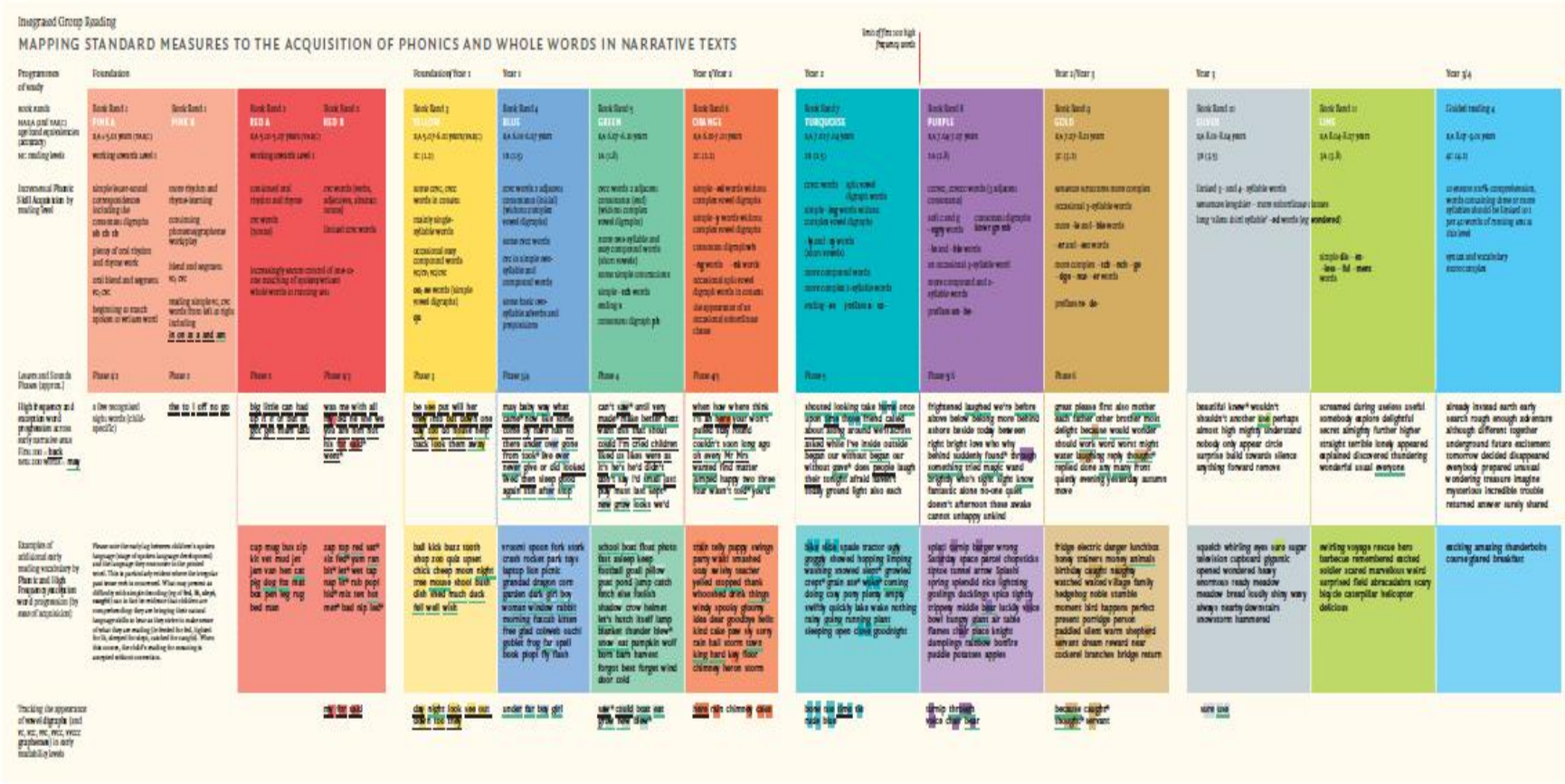
What Works Clearinghouse (2013). *Updated intervention report: Reading Recovery*. U.S. Department of Education, Institute of Education Sciences.

Wyse, D., & Goswami, U. (2008). Synthetic phonics and the teaching of reading. *British Educational Research Journal*, 34(6), 691–710.

Wyse, D., & Styles, M. (2007). Synthetic phonics and the teaching of reading: the debate surrounding England's 'Rose Report'. *Literacy*, 41(1), 35-42.

Appendices

1. Mapping Standard Measures to the Acquisition of Phonics



Integrated Group Reading: a contribution to classroom-based differentiated reading teaching

Time	IGR lesson checklist	Rationale*
1’	Briefly acknowledge children’s drawings and sentences from their TA session	Supports oral language and comprehension Builds links to TA session, shows continuity of IGR teaching
6’	Briefly check recall and comprehension while you deal GoFish cards and play the GoFish game	Memory consolidation (Clay, 1994, p. 55) GoFish: overlearning at the level of the sentence
4’	Introducing new material with storytelling	Storytelling for deep engagement, interest and comprehension
1’	Give the children time to look freely through their own copies of the book	To secure interest and engagement – lack of distractions
5’	Lotto game for phonological-visual mapping	Receptive and expressive vocabulary Gives pupils a chance to get used to unfamiliar words before collaborative reading
8’	Collaborative reading and problem-solving	Models the pace and rhythm of reading Supports fluency and simultaneous comprehension Develops individual and collaborative problem-solving skills
1’	1-2 comprehension questions and links to TA session	Develops memory and reflection Builds links to TA session, shows continuity of IGR teaching
4’	SWAP phonics game	Phonics (well research-evidenced)
Just after the session	Complete the Daily record	The importance of detailed observation of and response to children’s reading behaviour (Clay, 1994)

**For further evidence from the literature, see section 1b in the report*

Integrated Group Reading (IGR) methodology, together with its accompanying learning materials, has been developed to replace guided or reciprocal reading methodologies for reading-delayed Year 2 and Year 3 children in the classroom until such time as they have become confident and competent early readers and can access higher level reading and comprehension work. It can also be used (more slowly) with Year 1 children who are ready for small group reading work.

In Key Stage One in particular, the analytic or metacognitive approaches of guided or reciprocal reading comprehension lessons with their emphasis on discussion *about* books or texts on the one hand, and the phonics ‘first fast and only’ approach to the acquisition of reading skill on the other, can leave some children ill-equipped for either task. This can also be the case for any child needing a supplementary, story-based, intermediate introduction to the practice and development of reading itself.

The Integrated Group Reading (IGR) approach could therefore be said to be a *pre*-guided or *pre*-reciprocal reading methodology for children who struggle with what is currently on offer in schools. It offers a distinct and sophisticated teaching approach for pupils who cannot be fully helped through other avenues, with the particular aim of engaging or re-engaging them both with reading and with a sense of themselves as confident, succeeding readers.

Clay, M. (1994). *Reading Recovery: A guidebook for teachers in training*. Portsmouth: Heinemann

2. Example from IGR programme materials

Developed by Jan Stebbing

3. Registration of RCT and IGR team

The IGR evaluation has been registered as a trial with ISRCTN registration number (ISRCTN3842799X).

The programme and evaluation arm teams were both based in the Graduate School of Education, University of Exeter, but operated separately. The programme arm team included 4 local literacy advisors in the project local areas. The evaluation team was advised by Professor Jane Hurry (UCL Institute of Education). Part of her role was to ensure the rigour of the evaluation methods used.

4. IGR programme costs

Table X: IGR programme costs

Total IGR Programme Materials costs (covering Nuffield and school paid elements and in Phases 1 and 2; also, any additional additions to packs)	**Phase One** £19,954.24	**Phase Two** £19,954.24
Initial Adviser Meeting (Bristol)	£1,864	nil
Initial training costs (2 national day conferences - covering venue, food travel etc.)	£10,280	£10,280
Midway local top up trainings (both Phases across 4/5 sites)	£3,060 (4 sites)	£3,555 (5 sites)
Advisor costs (daily rate, travel costs both Phases including TA Training, covering what we paid but also what was provided without costs to be estimated)	£12,800 (as per 16 schools, 32 classes)	£13,840 (as per 15 schools, 32 classes)
LA review day costs (4 sites both phases;	£3,752 (4 sites)	£4,312 (5 sites)
TOTAL	£51,770.24	£51,941.24
NUMBER OF SCHOOLS	16	15
COST PER SCHOOL (2 classes)	£3,231.19	£3,246.33
NUMBER OF IGR CLASSES	2	2
COST PER IGR CLASS	**£1,615.59***	**£1,623.16***

**An additional early top-up training for each Local Authority adds another £3,060 to the total – just under £100 per class for four Local Authorities*.

The cost analysis above presents the costs associated with the arrangements for the IGR evaluation, and one could expect that the actual costs of using the programme beyond the evaluation would be slightly different. It is difficult to estimate these with accuracy here. IGR involves a one-off cost of materials and training for teachers and teaching assistants, with this cost being about £1,600. Subsequent years of implementation should not incur any materials/ training costs, as the same set of materials can be re-used and teachers can train other teachers and TAs in using the programme.

However, the programme has particular staffing requirements, and in order to be implemented in a smooth way a regular teaching assistant should be available to deliver the consolidation IGR sessions and work with the rest of the class when the teacher delivers IGR. Given the current school budget restrictions, this is a matter that schools should take into consideration when planning for IGR.

5. Table with acceptable and less acceptable variations to the programme

Variations that under certain conditions *can be acceptable*	**Originally expected practice**	**Notes**
Pace		
1. 1 book per week	2 books per week	Teachers should work towards the goal of covering 2 books
Class organisation		
2. During the teacher-led session, teacher and TA work together supporting 2 pupils each	During the teacher-led session, the teacher supports all pupils	Depending on the number of available TAs and the pupils' needs. Note that if the TA is working alongside the teacher during the teacher-led session, then this is in addition to, rather than instead of, the TA-led support sessions.
Routine		
3. Steps in the sequence are left out	Each session follows the sequence fully	This should only happen at an early stage of learning to use the IGR routine
4. Steps in the sequence are moved to the next session – e.g. the phonics game	No steps are moved to the next session	See comment above
5. Previous book recap and new book introduction are rushed	All steps are getting enough attention	All steps should get enough attention as they supplement each other
Strategies		
6. The back cover text is read silently	The back cover text is read aloud	This should only happen at an early stage of learning to use the IGR routine
7. Collaborative reading **begins** with the back cover being read in unison	Collaborative reading **ends** with the back cover being read in unison	It should be avoided for all pupils as the story context and the pictures can support their reading
Materials		
8. New resources made for special cases (e.g. for religious reasons)	To use the original resources	When special conditions apply
9. The cards in the games are open for all to read them	The cards in the games are protected to foster interest	Only till the pupils are used to play the games
Context		
10. Gaps in implementing IGR (before and after school holidays, teacher or pupil absence)	IGR is consistently implemented	When special conditions apply

Variations that depart from the IGR methodology and *should be avoided*	**Expected practice**	**Rationale**
1. 1 teacher-led session plus 3 TA-led support sessions	2 teacher-led plus 2 TA-led support sessions	The significance of the IGR programme is that the teacher is responsible for leading the IGR teaching and the TA-led support sessions follow and supplement the teacher-led sessions
2. TA introduces a new book	Teacher introduces a new book	As the teacher has the central teaching role in the IGR group she or he should introduce the new books
3. In the place of the phonics game, a phonics task tailored to the statutory phonics test	To use the IGR phonics game	The integrity of the IGR methodology needs to be kept during the intervention. Phonics is in IGR, and any additional remedial task can be done outside of IGR.
4. The Lotto game is played **after** the new book reading	The Lotto game is played **before** the new book reading	The IGR routine is both research-based and practice informed and should be followed in the right order. The Lotto introduces pupils to the new vocabulary before reading the book for pupils to become familiar with the perception of an unknown word. As such, it should always precede the new book reading.
5. Adding written comprehension task to the original routine	To keep to the current steps which include oral comprehension tasks	Reading comprehension is covered by IGR. Adding a written comprehension task can add an extra barrier to struggling pupils. The project is monitoring pupils' responses to reading comprehension.
6. Lotto game is played as a matching rather than a recognition activity	Lotto is played as a recognition game	Lotto as a matching game should only happen at an early stage of learning to use the IGR routine. With Lotto as Recognition the teacher cues the children by saying 'Everyone ready to Listen', and then *says but does not show them each card* until they have 'found and covered' the word they have just heard. This helps with aural-visual identification i.e. it gives the teacher a deep level of diagnostic information re the children's aural and visual perceptions before they read.
7. The TA repeats the teacher-led lesson	TAs have a different and supplementary role	The TA role is different to that of the teacher. TA-led support sessions aim to consolidate the new knowledge that the teacher has previously introduced.
8. A story reading is done as reading round rather than collaborative reading	Collaborative reading is an orchestrated mix of choral and individual reading	Collaborative reading should involve 2 elements: 1) A mixture of choral and individual reading, and 2) Collaborative problem solving, in the sense that each pupil can ask for the help of their peers. The teacher can facilitate this by asking 'can anyone help with this' or 'Knock for a Neighbour'
9. Go Fish game is used in the same session as the book it was based on	Go Fish should be played in the next session from the one that introduced the book that the game is based on	Go Fish is an important opportunity to recall previous knowledge and serves as a recap of the previous lesson
10. The rhyme is omitted when working with the Wizard Books	When working with Wizard books the lesson begins with a Read-Together of the Rhyme, and ends with a remembering/revisiting of the Rhyme, again as a Read-Together activity	The ability to rhyme is an important stage in children's phonological development that helps underpin their development as readers. The opportunity to practice rhyming in a supported way is integral to the research-based and practice-informed IGR routine

6. **Teacher-completed pupil identification form**

Teacher rating form to identify the pupils who will get the IGR programme intervention as part of the same reading group for 2016-17

1. In your class identify up to 10 pupils with the lowest reading attainment scores who might benefit from intensive IGR teaching.
2. Write their names in the table below.
3. Based on your knowledge about their relative attainments, give them a rating from 1 to 4 by choosing the appropriate box below.
4. Repeat for all 10 pupils.

TABLE 1. The student's reading level is:

	Children's names	**Very well below average in comparison to their peers** 1	**Well below average in comparison to their peers** 2	**About average in comparison to their peers** 3	**Above average in comparison to their peers** 4	**If score is 1 OR 2 put a tick in this column**
1		1	2	3	4	
2		1	2	3	4	
3		1	2	3	4	
4		1	2	3	4	
5		1	2	3	4	
6		1	2	3	4	
7		1	2	3	4	
8		1	2	3	4	
9		1	2	3	4	
10		1	2	3	4	

5. If the pupil scored 3 or 4, do not continue further. If they scored 1 or 2, write their names in the table below and identify the area(s) of difficulty that apply to each child by rating them on the following scale: 1. **not at all**, 2. **slightly**, 3. **very much.**
6. If more than 6 students score 1 or 2, please make a professional judgment as explained below (step 11)
7. **Then add up the scores for each child and insert in right hand column.**
8. Give a score for all areas of reading difficulty.

TABLE 2. Areas of reading difficulty for each child with scores 1 or 2 from previous table (table 1)

1. Not at all, 2. Slightly, 3. Very much

	Child's name from table 1	Difficulty with **Decoding**	Difficulty with **Fluency**	Difficulty with **Vocabulary**	Difficulty with **Comprehension**	Difficulty with **Motivation**	**TOTAL SCORE**	**TICK IF ONE OF THE FOUR WITH HIGHEST SCORES**
1		1 2 3	1 2 3	1 2 3	1 2 3	1 2 3		
2		1 2 3	1 2 3	1 2 3	1 2 3	1 2 3		
3		1 2 3	1 2 3	1 2 3	1 2 3	1 2 3		
4		1 2 3	1 2 3	1 2 3	1 2 3	1 2 3		
5		1 2 3	1 2 3	1 2 3	1 2 3	1 2 3		
6		1 2 3	1 2 3	1 2 3	1 2 3	1 2 3		

9. The **4 pupils with the highest scores** will be in the IGR group.

Please note

10. The 4 pupils in the IGR group will be **part of the same reading group** throughout 2016-17, and **the group would have to remain unchanged.**
11. If some children have the same score, **please make a professional judgment**, and the pupils left out of the group will be part of another group and could receive the IGR programme materials but not to the same extent as the IGR group.
12. **If you have difficulties with the selection**, or you are concerned that some pupils might not be able to access the programme materials please contact the project team.

7. Power analysis

Before the beginning of the study we conducted a power analysis to determine the required sample size. Statistical power is our ability to identify treatment effects as statistically significant at a certain level with a sample of a given size. In a simple randomised trial, to identify an effect of size 0.5 (medium size effect) as statistically significant at the 95% level with an 80% probability using a two-sample t-test we would need a sample of size 64 in each group (treatment and control). For an effect of size 0.4, we would need 99 observations in each group. These calculations were performed with the *pwr* package in R.

We had a clustered design and needed to account for its effects while estimating power. The correspondence between the number of observations required in a simple randomised trial and a cluster randomised trial is given by a formula (Campbell and Walters, 2014):

nCRT = nRCT * DE,

where nCRT is the number of subjects in a cluster randomised trial (CRT) in each arm, nRCT is the number of subjects in a simple randomised controlled trial (RCT) in each arm and DE is the design effect:

DE = 1 + (m - 1) * ICC,

where m is the number of subjects in each cluster and ICC is the intra-class correlation, i.e. the proportion of the total outcome variance that can be accounted for by between-cluster variance, or in other words, the expected correlation between two randomly selected subjects in the same cluster.

In our study we adjusted outcome measures for baseline measures of the same variables (the pre-test-post-test design) and other covariates. This will reduce the total outcome variance and increase power. After adjustment, the required number of subjects is:

nCRT = nRCT * DE * (1 – r2),

where r is the correlation between the baseline and post-treatment outcome measures or, more generally, r2 is the R-squared in a multiple regression model that regresses post-treatment outcome measures on baseline measures and other covariates, but does not include the treatment status.

We clustered pupils at the school level, with on average about 8 pupils per school (m = 8)[1]. We assumed that the ICC for the Hodder test at the school level is 0.1. The correlation coefficient between Hodder scores at time 1 and 2 was assumed to be 0.5. Hence, for the effect of size 0.5 we would need64 * (1 + (8 – 1) * 0.1) * (1 – 0.5^2) = 82 observations in each treatment and control group (i.e. about 10 schools in each group).

For the effect of size 0.4 we would need 126 (about 15 schools) observations in each group.

After the data collection we have got between 110 and 120 pupils in each group with non-missing data on the outcomes. Therefore, we were able to identify medium size effects with the power of 0.8, but not smaller effects.

[1] Here we report the power calculations that we did before the study. When doing the statistical analysis, we clustered the observations at the level of IGR group (usually 4 pupils) / class rather than school. When m is smaller the required sample size is smaller too, so the calculations in this section represent a more conservative estimate.

8. Statistical analysis methods

We considered five outcome variables for the IGR group analysis: the single word reading test score, Hodder test score, attitude to school scale, reading self-competence and attitude to reading scales. For the non-IGR pupils, we only considered the Hodder test scores.

In all statistical analyses, we applied the following procedure. First, for each pupil we calculated the difference between the values of the outcome variables before and after the IGR. Then we tested whether the pre- and post-treatment difference in the outcomes was statistically significantly different in the treatment and control groups by regressing it on the treatment status variable, controlling for gender, Year Group, special educational needs (SEN) status and for having English as additional language (EAL).

Given the clustered design of the study, we needed to correct standard errors for within-cluster correlation. This can be done either by applying cluster-robust standard errors to the linear regression (as in the survey R package) or by fitting multilevel models (using the lme4 R package). We applied both methods, and the results are similar. In this report we use the models with cluster-robust standard errors. For the IGR group analysis, the observations were clustered at the level of an IGR group. For the rest of the class analysis, the observations were clustered at the class level.

9. Ethics

The project had ethical clearance from the University of Exeter, ensuring the voluntary and informed nature of participation, special arrangements for equal access for participants (none required), assessment of possible harm (no harmful effects were involved) and data protection procedures. There was particular reference to the way the programme and evaluation teams operated, as they were both based at the University of Exeter but worked independently from each other to ensure that there was no bias in the evaluation practices.

All participating schools signed a memo of understanding outlining the project's procedures and a consent form. Informed passive consent was sought from parents, for both pupils in the identified groups and class pupils, and letters were sent explaining what the randomisation process involved (that the comparison pupils would be on a waiting list for the programme) and were distributed before the randomisation took place. Some schools requested an extra consent form to be produced for the collection of the demographic data for the participating class and pupils. Anonymity and confidentiality has been applied to every aspect of the project, and school/ individual participants had the right to withdraw at any time. In order to not affect the evaluation, we did not communicate the scores from the assessments to the schools. These will be sent to them after the end of phase 2 (September 2017) as aggregated results, and to individual school by request only – schools will be able to access their own individual results only.

10. Detailed statistical findings

Participants

As we can see from Table 1, for most of the demographic characteristics there was no statistically significant difference between the treatment and control groups. The exception was the percentage of children with SEN on school support. There were more children with SEN in the control schools compared to the treatment schools in phase 1; this might suggest that in phase 1 some treatment schools could perceive IGR as a substitute to SEN School Support provision and so not identify pupils having IGR as SEN Support. Also note a much higher proportion of missing values for the phonics score in the control schools.

The effects on IGR pupils

We present the statistical results in the tables below. The IGR effect was calculated as the coefficient for the treatment status in the model with cluster-robust standard errors (at the IGR group level) that control for year, gender, SEN and English as additional language status. The p-value is the p-value for these coefficients from the same model. Cohen's d is the standardised effect size (IGR effect divided by the standard deviation of the outcome).

As can be seen from the tables 6 and 7, none of the IGR effects for any of the outcomes in either Phase 1 or 2 are statistically significant at the 90% or 95% level. The effect sizes are mostly close to zero and are never larger than 0.25. Thus, we do not have enough statistical evidence to conclude that IGR had a positive or negative effect on IGR pupils' outcomes compared to the alternative programmes used in the control classrooms.

Both in the treatment and control groups, pupils showed progress on the standardised reading test scores between the pre-treatment and post-treatment measurements. We summarise the progress in the tables below.

Table 5. IGR: IGR and comparison groups

	Treatment phase 1 (N=131)	**Treatment phase 2 (n = 126)**	**Control phase 1** (N=132)	Missing in the treatment phase 1 group (%)	Missing in the treatment phase 2 group (%)	Missing in the control group (%)
Boys	84 (64%)	70 (56%)	89 (67%)	0 (0%)	0 (0%)	0 (0%)
Year 2	63 (48%)	63 (50%)	71 (54%)	0 (0%)	0 (0%)	0 (0%)
Non-English ethnic background	25 (19%)	34 (27%)	23 (17%)	22 (17%)	7 (6%)	18 (14%)
English as an additional language	19 (14%)	25 (20%)	18 (14%)	22 (17%)	7 (6%)	18 (14%)
SEN (Education, health and Care Plan (EHC Plan)	5 (4%)	6 (5%)	7 (5%)	22 (17%)	7 (6%)	18 (14%)
SEN (school support) *	38 (29%)	49 (39%)	63 (48%)	22 (17%)	7 (6%)	18 (14%)
Pupil Premium	30 (23%)	37 (29%)	34 (26%)	22 (17%)	7 (6%)	18 (14%)

Child in Care	0 (0%)	1 (1%)	4 (3%)	22 (17%)	7 (6%)	18 (14%)
Mean phonics score **(40 is the maximum)**	24.0	28.4	27.4	36 (27%)	37 (29%)	61 46%)

Note: The variables where the difference between the treatment and control groups is statistically significant at the 95% level are marked with an asterisk (*).

Table 6. Phase 1 results: IGR pupils

	Control Time 1	Control Time 2	Control n	Treatment Time 1	Treatment Time 2	Treatment n	Control T2 - T1	Treatment T2 - T1	IGR effect	Cohen's d	p value
HGRT standard score	86.1	90.3	117	90.5	92.2	112	4.18	1.77	-2.42	-0.23	0.20
SWRT standard score	85.6	89.6	118	86.3	89.5	118	4.03	3.10	0.13	0.01	0.93
HIFAMS: attitude to school	1.6	1.6	118	1.7	1.6	118	-0.07	-0.04	0.02	0.07	0.63
HIFAR: reading self-competence	3.4	3.5	117	3.4	3.5	116	0.11	0.13	0.05	0.06	0.68
HIFAR: reading attitude	3.9	4.0	117	3.9	4.2	116	0.14	0.22	0.10	0.12	0.42

Table 7. Phase 2 results: IGR pupils

	Control Time 1	Control Time 2	Control n	Treatment Time 1	Treatment Time 2	Treatment n	Control T2 - T1	Treatment T2 - T1	IGR effect	Cohen's d	p value
HGRT standard score	86.1	90.3	117	90.2	96.2	119	4.18	6.01	2.49	0.24	0.15
SWRT standard score	85.6	89.6	118	89.2	92.7	118	4.03	3.46	-0.05	-0.01	0.96
HIFAMS: attitude to school	1.6	1.6	118	1.6	1.6	118	-0.07	-0.01	0.06	0.16	0.23
HIFAR: reading self-competence	3.4	3.5	117	3.3	3.5	118	0.11	0.16	0.09	0.11	0.42
HIFAR: reading attitude	3.9	4.0	117	3.9	4.0	118	0.14	0.17	0.05	0.06	0.69

Table 8. Phase 1 results: reading ages (years: months)

	Control Time 1	Control Time 2	Control n	Treatment Time 1	Treatment Time 2	Treatment n	Control T2 - T1	Ratio gains	Treatment T2 - T1	Ratio gains	IGR effect
HGRT	5:7	6:7	117	5:10	6:9	112	0:11	1.5	0:11	1.5	0:0
SWRT	6:1	7:0	118	6:1	7:0	118	0:11	1.5	0:11	1.5	0:0
HGRT: non-IGR pupils	7:3	8:5	573	7:4	8:8	586	1:2	2	1:4	2.2	0:1

Table 9. Phase 2 results: reading ages (years; months)

	Control Time 1	Control Time 2	Control n	Treatment Time 1	Treatment Time 2	Treatment n	Control T2 - T1	Ratio gains	Treatment T2 - T1	Ratio gains	IGR effe
HGRT	5:8	6:7	117	5:10	7:0	119	0:12	1.7	1:2	2	0:3
SWRT	6:1	7:0	118	6:3	7:1	117	0:11	1.5	0:10	1.4	-0:2
HGRT: non-IGR pupils	7:4	8:6	573	7:3	8:6	598	1:2	2	1:3	2.1	0:1

IGR: interaction effects

We have also explored the interaction between the IGR and a number of socio-demographic characteristics of pupils, in order to test the hypothesis that the IGR effect can be different for boys and girls, pupils in Years 2 and 3, native and non-native English speakers, with and without special education needs or Pupil Premium status. Note that even for the identification of the main effects we only have the power to identify medium-size effects. For interactions, the power is even smaller.
The results of these analyses are reported in the Appendix 15.

Overall, we do not find consistently statistically significant interaction effects between the IGR and any of the socio-demographic variables. In phase 1, IGR girls did significantly better on the Hodder test scores than IGR boys, but this effect was not replicated in phase 2 nor with the SWRT test in either phase. Similarly, in phase 1 IGR children with English as an additional language showed better progress compared to native speakers, but again the effect was not replicated in phase 2 or for SWRT scores. In phase 2 (but not phase 1), IGR pupils with the Pupil Premium status did significantly better on the SWRT test (but not on HGRT).

Given a large number of hypotheses we are testing for the interaction effects, some of these effects are likely to be statistically significant just by chance. None of the effects is consistent for both measures of the reading progress and in both phases of the study.

IGR: long-term effects

We have also explored the long-term IGR effects for Phase 1 children. The table below summarises the results. We report the results only for the children for whom both time 1 and time 4 measures were available. Time 4 refers to a time period of two years after the start of the IGR intervention (phase1 implementation) and 9-10 months after the programme evaluation was ended (IGR might have been used in some form or extent or not).

The only statistically significant long-term IGR effect is for the Hodder scores. The progress on the Hodder scores between times 1 and 4 was larger in the control schools compared to the treatment schools, and the IGR effect is negative. The effect size is approximately -0.3. Note, however, that at time 4 the mean Hodder scores in the control and treatment groups were very similar. The negative IGR effect is due to the lower Hodder scores in the control group at time 1 compared to the treatment group. It is possible that the measurement error for the Hodder scores at time 1 affected these results. The long-term IGR effect on the non-IGR children (table 10) is positive, but small and not statistically significant.

Table 10. Long-term IGR effects

	Control Time 1	Control Time 4	Control n	Treatment Time 1	Treatment Time 4	Treatment n	Control T4 - T1	Treatment T4 - T1	IGR effect	Cohen's d	p value
HGRT standard score	85.6	90.2	106	90.2	90.8	110	4.62	0.55	-3.72	-0.32	0.05
SWRT standard score	85.4	89.1	108	86.2	89.1	107	3.70	2.93	0.39	0.04	0.80
HIFAMS: attitude to school	1.6	1.6	108	1.6	1.6	107	-0.05	-0.06	-0.02	-0.06	0.72
HIFAR: reading self-competence	3.4	3.4	107	3.4	3.4	105	-0.06	0.01	0.02	0.02	0.89
HIFAR: reading attitude	3.9	3.8	107	3.9	4.0	105	-0.07	0.10	0.18	0.18	0.24
HGRT standard score: non-IGR pupils	106	107	520	105	108	548	0.73	2.8	1.8	0.15	0.21

The IGR effect on the non-IGR pupils (rest of the class)

We can also test if the IGR affected the outcomes of the pupils who were not directly involved in the IGR programme (the rest of the class). These children could be affected because of the change in teaching routines in the class, change in teacher's behavior, etc. Since the sample size for these children is larger, the analysis has more statistical power and we should be able to identify smaller effects. For this analysis we only have Hodder scores as the outcome, and we are unable to control for the SEN and EAL status as this information was not collected.

Table 11. Phase 1 results: non-IGR pupils

	Control Time 1	Control Time 2	Control n	Treatment Time 1	Treatment Time 2	Treatment n	Control T2 - T1	Treatment T2 - T1	IGR effect	Cohen's d	p value
HGRT standard score	106	108	573	105	108	586	1.7	3.3	1.7	0.13	0.24

We see that in both phases the IGR showed a small positive effect on the reading progress (as measured by the Hodder scores) of the non-IGR children. The standardised effect size was 0.13 in Phase 1 and 0.23 in Phase 2. In phase 2 the effect was statistically significant at the 95% significance level (p = 0.03).

Table 12. Phase 2 results: non-IGR pupils

	Control Time 1	Control Time 2	Control n	Treatment Time 1	Treatment Time 2	Treatment n	Control T2 - T1	Treatment T2 - T1	IGR effect	Cohen's d	p value
HGRT standard score	106	108	573	105	109	598	1.7	4.5	2.8	0.23	0.03

However, we should be careful with not over-interpreting these results. When we fitted the interaction effect between the IGR and gender in the models for the non-IGR pupils, the positive IGR effect was only present for girls. The explanation for this is significantly higher pre-treatment baseline mean Hodder scores for girls in the control group. The table below illustrates these findings.

Table 13. Gains in HGRT for non-IGR pupils in treatment and control groups for boys and girls

TreatmentStatus	gender	baseHodder	afterHodder	diffHodder	n
Control	Boys	104	108	4.20	259
Control	Girls	108	107	-0.56	302
Treat	Boys	105	108	3.28	292
Treat	Girls	105	108	3.22	292
treat phase 2	Boys	104	108	4.39	313
treat phase 2	Girls	106	110	4.64	284

In all the conditions, except girls in the control group, pupils showed approximately the same progress in the Hodder scores between the pre- and post-treatment measures (between 3 and 5 points). Only the girls in the control group did not show any progress. If we take these findings at face value this means that the traditional organisation of the classroom benefited boys, but not girls, and that IGR corrected this gender disparity. Another explanation is that a measurement error for the girls in the control group at time 1 introduced a possible upward bias. We have also explored the interaction effects between the IGR and school year; these are not statistically significant.

Fidelity

Table 15. IGR fidelity – sequence

	No of IGR pupils		Mean gain in IGR pupils' SWRT		Mean gain in IGR pupils' HGRT		Rest of class mean gain in HGRT	
Sequence followed	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2
Not full & poor quality	4	0	-4.3***	-	-5.7***	-	9.6***	-
Some extent & moderate quality	43	35	3.4	3.2	0.8	6.3	4.0	6.7**
Fully & good quality (reference group)	68	84	3.5	3.6	2.4	6.0	2.5	3.8

Note: Stars indicate statistical significance of the difference with the reference group. *** p <0.01, ** p<0.05, * p<0.1. Models with clustered standard errors were used to estimate statistical significance.

Table 16. IGR fidelity – number of sessions

	No of IGR pupils		SWRT mean IGR pupils' gain scores		HGRT mean IGR pupils' gain scores		Rest of class mean gain in HGRT	
Number of sessions per fortnight	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2
2-5	20	6	2.6	-3.0	0.6	8.7	2.0	-7.3***
6-8 (reference group)	111	82	3.2	3.0	2.0	6.2	3.5	5.7

Note: Stars indicate statistical significance of the difference with the reference group. *** p <0.01, ** p<0.05, * p<0.1. Models with clustered standard errors were used to estimate statistical significance.

Table 17. IGR fidelity – used with other pupils

	No of IGR pupils		SWRT mean IGR pupils' gain scores		HGRT mean IGR pupils' gain scores		Rest of class mean gain in HGRT	
IGR used with other pupils	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2
No (reference group)	75	67	1.6	2.5	-0.6	6.7	2.4	4.9
Sometimes	12	26	4.3	5.0	3.3	2.4	6.8*	2.6**
Yes	36	22	5.8	5.7	6.5*	9.4	3.4	7.2

Note: Stars indicate statistical significance of the difference with the reference group. *** p <0.01, ** p<0.05, * p<0.1. Models with clustered standard errors were used to estimate statistical significance.

Table 18. IGR fidelity – delivery location

	No of IGR pupils		SWRT mean IGR pupils' gain scores		HGRT mean IGR pupils' gain scores		Rest of class mean gain in HGRT	
Take IGR pupils out of class	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2	Phase 1	Phase 2
Out	20	4	3.7	7.5***	1.7	22.7***	2.6	12.0***
Sometimes out	12	4	-0.1	1.5*	3.0	0.0***	2.9	2.9
In class (reference group)	95	107	3.6	3.6	1.4	6.1	3.6	4.4

Note: Stars indicate statistical significance of the difference with the reference group. *** p <0.01, ** p<0.05, * p<0.1. Models with clustered standard errors were used to estimate statistical significance.

11. Process evaluation methods

Contexts Mechanisms Outcomes analysis

14 schools (8 in phase 1 and 6 in phase 2) (mixed range of rural, sub/urban schools), each acting as different cases, were selected across the 4 LEAs based on a combination of differing levels of a school characteristic (percentage of pupils receiving Free School Meals) and a teacher characteristic (different levels of teacher self-efficacy for teaching reading) – the latter was based on a reading teaching self-efficacy questionnaire completed by teachers as part of the training day, which will be further discussed later on.

The process evaluation was designed using a realist evaluation framework (Pawson & Tilley 1997). An IGR programme theory was constructed in terms of over 100 context, mechanisms and outcome areas expressed in both positive and negative terms to inform the data collection and analysis.

In each school one (or more) teacher-led IGR session/s was observed and one (or more) teacher/s was interviewed and completed an IGR context, mechanism and outcomes (CMO) questionnaire based on a shorter version of the programme theory. The questionnaire asked for the perceived interim outcomes of the programme (10 items) and the mechanism (20 items) and context factors (10 items) that had supported them (appendix 18). After each lesson observation, pupils in the intervention groups (IGR groups) were asked about their experiences of the programme.

Implementation data

These data were complemented by 10 online logs per phase (20 logs in total) completed every fortnight by all the IGR teachers to monitor how they implemented the programme. The log asked teachers to summarise their classroom organisation, the number of teacher and TA sessions, the programme's teaching routine, and comment on pupil attainment and attitude. It was reviewed and revised several times during the year to better capture departures from the suggested organisation and methodology. The log was radically redesigned in phase 2 to be shorter and easier to complete in response to feedback from phase 1.

In addition to the log, there were observations from the programme team support visits to schools (conducted at various intervals during the year and recorded on observation forms) and focus group interviews from local review meetings with the teachers at the end of two terms of the IGR programme – local review meetings were organised at the end of both phases.

Table 4 summarises kinds of data collected.

Table 4: Process evaluation data overview for both phases 1 and 2

Process evaluation data overview		
	Phase 1	**Phase 2**
Teacher-led IGR observations	10	7
Teacher individual interviews	12	14
CMO teacher questionnaires	12	13
Review meetings interviews	26 teachers in 4 focus group interviews	31 teachers in 5 focus group interviews
IGR pupil brief interviews	10	7

Programme team visits	observation forms, emails etc.	observation forms, emails etc.
The IGR log	fortnightly log (10 logs)	fortnightly log (10 logs)
Other	photos, emails, notes etc.	photos, emails, notes etc.

Control teaching

Control teaching was not explored directly through school visits to avoid putting additional pressure on control schools who were on a waiting list to use the intervention. An online survey was sent in November 2015 to the teachers who were in the control group and were continuing with their usual teaching. The survey asked for information about the teaching of reading and the classroom organisation. A second survey was sent at the end of phase 1 (autumn 2016), focusing more on the 4 identified comparison pupils and the use of additional literacy programmes. Two control teachers participated in a follow-up phone interview to discuss their responses in more depth. This applies only to phase 1 as there was no control group in phase 2.

Data analysis

School visit interviews were analysed in a grounded theory-style approach at a case level and then again across cases. Review meeting interviews were analysed in the same way but separately from the individual cases. The same thematic structure informed by the programme theory was used using the NVivo 11 programme. Information from the log and the programme team observations was organised in themes by teacher. A summary for each school was initially produced (to synthesise school visit data, and data from the log and the programme team observations), and then a cross case summary was prepared to summarise common themes. This procedure was followed for both phase 1 and 2 data. Control teaching data (applicable to phase 1 only) were analysed separately and a summary was produced.

Fidelity data (fidelity refers to the quality of programme implementation)

A fidelity index was designed using information collected by the fortnightly log, and from observation data (from the process evaluation and the programme team support visits).

The index was a combination of scores on four dimensions: (1) the quantity (whether all steps in the IGR lesson routine were present and being delivered in the right sequence) and quality of the teaching in the sequence of IGR methodology, (2) the number of teacher- and TA-led sessions, (3) where IGR took place (in or out of the regular classroom), and (4) whether IGR was used with other pupils (outside of the IGR group as evidence of dedication to the programme). Each teacher delivering IGR was rated on all dimensions – where there was enough information – and a separate analysis was conducted for each dimension. This procedure was revised in phase 2 to better capture the fidelity of teaching. The other three dimensions were calculated in the same way across phases.

Phase 1: For the quantity and quality of sequence delivery, scores were given as follows: 0: not full and of low quality; 1: sequence evident to some extent and of moderate quality; and 2: sequence evident in full and of good quality. In phase 1, this was based on observations from both the programme and evaluation arms, with scoring from both arms taking account of whether the expected sequence of IGR lesson was evident and of good quality. Ratings from the evaluation and programme team were compared to ensure the validity of this rating system (observations were conducted independently and using different observation guides). 9 of the 31 teacher ratings from both sets of observations were available for comparison, and of these, 6 out of 9 teachers had equivalent ratings, with 3 out of 9 showing a difference of 1 point only and none had a 2-point difference. Therefore, for 9 out of 9 teachers the rating was the same or within 1 point of equivalency, suggesting that programme and evaluation arm teacher observations were moderately consistent.

Phase 2: A new index was devised (appendix 19) that allowed for a more detailed scoring of the phase 2 teaching using the 3-point scoring system described above. The index though could not be used to recalculate the phase 1 fidelity because the data available for phase 1 was not sufficient for a more detailed analysis. Programme team records were more systematic in phase 2 visits. Based on the programme team phase 2 observations and using the

fidelity index, the programme and evaluation team scores of the same teachers correlated highly (0.8), but the programme team scored all teachers consistently lower (a mean of 1.8 compared to 2.3) (The scores used in the analysis were the scores of the evaluation team).

Both phases: The number of teacher- and TA-led sessions was calculated from the log. The range was from 0-8 sessions per fortnight. The rating system was based on mean records from the online logs, namely up to 4 sessions for teachers and 4 sessions for TAs per fortnight.

The log also had a question about whether IGR was used with other pupils, outside of the IGR group (thus not monitored for the evaluation). This was considered to indicate the teachers' dedication to and interest in the programme. The following scores were used: No = 0; Sometimes = 1; Yes = 2. When the IGR programme was used with other pupils, this was mostly done in a loose way. This information was based on the log, but where there was missing data, observation data was used to complete the data set. In phase 1, for those teachers whose log entries were no or sometimes, the visit observations were in agreement with the log in all 11 cases compared.

Both phases: Re the location of IGR (in or out of the regular class) the following scores were used based on the log: Not in class = 0; Sometimes not in class= 1; In-class = 2. In phase 1, 20 out of 21 teachers who reported that IGR took place in the class were also observed to be doing this. 4 out of 5 teachers who reported that IGR happened out or sometimes out of the class were also observed doing this.

12. Detailed process evaluation findings

Setting up the programme implementation

School recruitment

The difficulties in recruiting schools for the project have been discussed earlier (4.b.i) and are revealing of the pressures that the RCT design can put on schools, namely the staffing and organisational demands, the waiting list design for control schools, the tight assessment schedules, and the overall lack of flexibility.

Pupil identification

As discussed, before the beginning of phase 1, teachers were asked to identify four pupils who would benefit from literacy support and were given a form and specific instructions to help them with the selection (adapted from Speece et al., 2011) in accordance with the protocol of the study. Details of the approach are discussed in the methods section (4.b.ii).

Yet, following this procedure a number of issues arose. In phase 1, as the identification of children had taken place in the summer term (summer 2015), some teachers had not been involved in the selection either because they were not starting their employment until the September, or because the teacher who had had the class at the time plus their school leaders had made the selection. Even when the selection happened in early September, some teachers did not yet have good knowledge of the children. As a result, not all teachers were happy with their IGR groups once the programme started and wanted to replace children who they felt were less suitable for IGR with pupils who would benefit more. Such changes were discouraged and kept to a minimum (7 instances) and new pupils were not monitored. In phase 2, the pupil selection took place at the end of summer 2016 and teachers were asked to involve as far as possible the Year 2 control pupils who would be in Year 3 in 2016-17, and to identify new Year 2 pupils (then in Year 1). During the training (July 2016), it was made clear that changes to the groups would have to be avoided, and despite the gap between the training (previous summer) and the phase 2 IGR implementation (October 2016), no teacher made changes to their groups after the phase 2 time 1 assessments (September 2016) – changes in the numbers of IGR pupils between the phase 2 time 1 and 2 assessment times were all due to pupils leaving school (N=3).

An issue that arose in both phases was how a teacher could deal with having pupils with different reading abilities in the IGR group. A discrepancy of abilities in some groups was partly because the reading progress of some children accelerated in response to IGR teaching, and partly because the initial identification was done in the summer term

and circumstances had changed. In some of these cases, the programme team suggested that the teacher could use different level books in different sessions to help ensure that all needs were met. Some teachers also preferred to make sure that IGR pupils who progressed ahead of the others had reading input relevant to their level in addition to the IGR programme, since their participation in the group was seen to help their self-confidence. However, in one case in different schools in each of phases 1 and 2, the programme team advised that the IGR group be divided into 2 sub-groups as they needed separate teaching using differently-levelled IGR materials. This solved the problem of instruction but created organisational difficulties as the teachers had to organise separately for each sub-group (and for both teacher/TA sessions).

Teachers were also encouraged to use IGR materials (following the programme's protocol or not) with other pupils in their classes and many phase 1 and 2 teachers did, although pupils in these groups were not monitored individually for the experimental evaluation.

Classroom organisation

The schools that acted as case studies in the process evaluation were very varied regarding their overall pupils' FSM percentages (from 8.6% to 52.9%) and their rural to sub/urban locations. Most had a strong dedication to literacy teaching that took various forms, from additional phonics support often delivered in pull-out sessions by TAs, to more sophisticated models of supporting reading using Guided Reading materials (Oxford Reading Tree etc.) or commercial programmes such as Accelerated Reader, Lexia, and ReadWriteInc. Combining some of this provision (e.g. ReadWriteInc) with IGR proved in some cases to be challenging, since the organisational and staffing requirements could be conflicting (as evident in the teacher interviews). This proved more problematic in phase 1 than in phase 2, partly because, building on experience from phase 1, the phase 2 training prepared the teachers more thoroughly for the programme's organisational requirements.

The IGR programme was designed to be delivered during guided reading group organisation and involved the class teacher teaching the intervention group for 30 minutes twice a week, while the TA worked with the rest of the class. This took place in the regular classroom during a time that was literacy-related for all (Guided Reading or other form of group reading), and not in pull-out sessions. This organisation was described as 'marginally more demanding' by one teacher, and this seemed to reflect the overall attitude of all interviewed teachers. Most teachers could see the value of keeping all the pupils in the classroom during the intervention, but there were 2-3 teachers who saw a tension between the inclusive aspect of IGR (keeping all pupils in the class) and the difficulty of maintaining concentration in a busy and lively class:

> Year 2 teacher*: '…that's a tricky one […] they cope fine with the model as it is, but I sometimes think: we could take them out'.*

Pull-out sessions were seen as the usual model of additional support by many teachers:

> Year 3 teacher: *'Normally I wouldn't spend half an hour just with one group. I would wander around, speak to other children as I've said before listen to other children read… So, I think IGR would work really well, if the group of children were taken out of the classroom to do the activities'.*

However, this perspective contrasts with the responses to the CMO questionnaire where both phase 1 and 2 teachers rated highly the inclusive organisation of IGR and generally agreed that one of the most important teacher-related outcomes of the programme was confidence in its class organisation. Many teachers also saw IGR as an opportunity to provide relevant teaching to all their pupils in the 'Quality First' setting (teacher interviews).

Giving the main role to the teacher means also that a TA has to be available to work with the rest of the class. Some schools had issues with TA availability, and many teachers (with the exception of 1-2 experienced and confident teachers) stressed how crucial it was to have a TA available to work with the rest of the class during the teacher-led IGR sessions (group review meetings).

The main issue schools and teachers had was with the number of teacher sessions when there were more than four reading groups in the classroom (as evident in teacher interviews). With four reading groups already in each class, this meant that teachers had to fit these four sessions into three days for the period of the intervention. An alternative was for the teacher to see two groups on the same day or split their time between these two groups with the support of a TA. Teachers came up with a variety of solutions to this issue (as evident in teacher interviews and the log) with the most common being the delivery of one of the two teacher-led IGR sessions in the classroom but during a school assembly, so in a sense the IGR pupils were not taken out of the class, but the rest of the pupils were not present. This was discouraged in phase 1 as a departure from the suggested organisation (although many teachers used it to varying degrees), but it was presented as a last organisational option in phase 2 (with only 2 schools using it consistently) (appendix 4). Other teachers who happened to be SENCOs used their SENCO time, asked their jobsharer to join them for one of the sessions, involved the school leaders (e.g. the Deputy Head), or extra TAs. On the other hand, a few experienced teachers were able to read with 2 groups on the same day (for 15 minutes with each or at different times in the day).

In some schools, it was made clear that all reading groups should have an equal entitlement to the teacher's time. This is an example from a school where both teachers stressed that the parents would like their children to have teacher input every week:

> Year 3 teacher: *'It was not going to be acceptable to our parents for any of our children to have their teacher input to guided reading removed from them, and it wouldn't matter that it wasn't going to be permanent, that it would only be this week, and they would be heard next week'.*

In terms of the organisation of the rest of the class during teacher-led IGR sessions, most teachers felt that the rest of the pupils were engaged during IGR, but that some time was required for the other pupils to learn to work in an independent and meaningful way. The observations indicated that classes were generally settled when the teacher was teaching the IGR group, although some were noisy and other pupils were often seeking the teachers' attention. This is also evident from the responses to the CMO questionnaire item 'the IGR organisation model results in an ordered class'. The item was rated positively but as the 2nd (phase 1) and 3rd (phase 2) lowest of 20 other mechanisms (Appendix 16 and 17), suggesting that for some teachers the organisation of the rest of the class during IGR sessions could be an issue.

Issues related to programme implementation

Materials and support

At the time of the evaluation, IGR materials started initially at red/yellow readability band (a reading age equivalency of roughly 5.07 years) and progressed to turquoise (approximately 7.01-7.04 years). At the programme support team visits, it was noted that the materials that the pupils were using (especially in Year 2) were often more difficult than the children needed, so additional materials were supplemented with specially made IGR packs. This and the teachers' reports led to the conclusion that more materials were needed to cover the range of pupils' reading levels at the lowest and highest book bands (vertically), but also within each band for pupils who plateau-ed (horizontally) (programme team notes, teacher interviews and the log). As more resources were needed, in some cases teachers had to produce some by themselves, but this was seen by them as a problem.

Pupils were very enthusiastic about the materials, although a few teachers (2-3) found them less engaging. Teachers in urban areas noted that the materials could have been more culturally diverse (teacher interviews). The children liked particularly the soothing character of the books, the rural lifestyle presented in the stories, and being able to follow the same characters in different adventures. They also admired the illustrations and enjoyed the games.

IGR currently has book series in 4 genres: Old Mrs Winterbottom books; Traditional tales; Fables; and Abracadabra books. Teachers across both phases noted that books from the Abracadabra series were more difficult to be used for storytelling, since their stories were more basic compared to the other series. In relation to this, a few teachers reported that they found the content of the books to be too basic for Year 3 pupils whom they felt would normally

read much more advanced books – however the language and vocabulary level was appropriate for their pupils, thus creating an interesting contradiction. This could indicate either that current IGR book content is more relevant to Year 2 pupils.

Pupil attitudes to IGR

Teachers reported that pupils were not concerned about being visible in a low attainment group (CMO questionnaire, appendix 9a & 9b), and that being in the IGR group was often seen as a privilege because of the out of the ordinary activities, materials and games. Some class pupils were also reported to be envious of the materials, particularly the games. This applies to both phases of implementation. There were only a couple of cases indicating otherwise, especially the case of a boy in phase 1 whose parents asked that he be removed from the IGR group as they felt it had an adverse effect on his self-confidence. The matter was reported by the programme team who attempted to resolve the issue with the Head teacher and the class teacher who was in her first year of teaching. In addition, 1-2 phase 2 teachers reported that a few pupils in their groups expressed the wish to progress and be placed in more advanced groups where they could read books more relevant to their age.

Many teachers reported that pupils did not see IGR as an intervention – this was partly because of the games and fun element in the activities, and partly because it took place in the classroom, as illustrated below by a phase 2 teacher:

> Year 3 teacher: '*The children in my group do not see IGR as an intervention – it is just their reading. So, the idea that they are not being withdrawn doesn't even occur to them, because they are part of the class, as they would be […] for any other guided reading type activity. They just see [themselves] as part of the normal classroom*'.

Teacher attitudes to IGR

IGR proved to be a demanding approach as far as teacher skills were concerned, since it adopts a multi-perspective approach to reading that can be seen as different from the current approach to early literacy that emphasises synthetic phonics and inference. Teachers had mixed views on this: some younger teachers who had been trained with a focus on phonics tended to alter the delivery of IGR slightly (such as the games) to be closer to a more phonics-driven instruction (e.g. a phase 1 teacher was observed using ReadWriteInc strategies in IGR sessions). In a similar way, the story-telling element of IGR for some teachers tended to be altered into a more inference-driven approach to text with teacher questions and pupil responses, in a teaching style closer to the Guided Reading approach that aims to make pupils independent readers. On the other hand, many teachers appreciated the simplicity of IGR that combined a variety of light touch approaches to re-engage pupils in reading.

Building on the phase 1 experience, the programme team placed a lot of emphasis on the importance of keeping the IGR lesson as simple as possible (avoiding heavy questioning, inference, writing tasks and grammatical comments) in the phase 2 training and support. Yet, many teachers, although they appreciated the importance of simplicity for the teaching of this particular group, were concerned about the requirements of the National Curriculum in relation to SATs and the end of the year assessments. As IGR was taking place 4 times a week (2 teacher and TA sessions), not all teachers could easily find time in addition to IGR to practice phonics, comprehension and writing. This created a lot of stress for a few teachers across phases, especially to younger or inexperienced teachers, and, in some cases, this led to replacing some IGR sessions with SATs preparation sessions. This seems to suggest that, although pupils in the IGR groups took simplified end of the year tests, many teachers felt that the programme did not prepare the pupils for the requirements of these tests, as IGR did not cover inferenced-based tasks, writing and synthetic phonics. This is because IGR is a remedial programme that aims to support very weak readers to re-engage with reading and thus operates as a nurture group with less focus on particular achievement goals.

Positive and negative outcomes

Most teachers across phases reported in both their interviews and CMO questionnaires positive outcomes for IGR pupils in reading attainment and attitude and oral language skills (appendix 9a & 9b). Especially for pupils with English as an additional language, teachers reported that IGR books and illustrations generated discussions developing pupils' vocabulary and social language.

This is an excerpt from the fortnightly log of a phase 1 Year 3 teacher towards the end of the year, where the teacher discusses the impact of the programme on a particular IGR child:

> *'James [not his real name] has expressed that he 'likes this Guided Reading but not the old Guided Reading'. He is now buying books at home and is able to read his football magazines. […] He will ask for help from others by sticking his thumbs up. James now reads with expression and understanding, he will explain what has happened in the story'.*

In the group review meetings, most teachers stressed that their IGR pupils had become more confident. Some pupils e.g. those who used to hide their books so they would not have to read at home, were very enthusiastic about the IGR stories and wanted to share them with their parents. Some teachers reported that pupils had also developed their social skills, through working as a group with strategies such as turn-taking and knocking for help (collaborative problem solving) – in a way similar to the log excerpt above (pupil asking for peer help by sticking his thumbs up).

However, some teachers noted that these outcomes either in attainment or attitude and social skills were more evident in the safe environment of the IGR group. In other words, the new skills were not yet transferred, and the teachers believed that more time (in terms of exposure to the programme) was needed for the IGR pupils to be able to use their new skills out of the IGR setting. This is illustrated by the excerpt below from a phase 1 review meeting focus group interview:

> *'They're confident within the IGR and within their coloured reading book, but when it comes to SATs tests […] it just didn't transmit, it was like a big block'. […] I think it is to do with the fact that the confidence is within that setting, and it's not quite transmitted outside of it'.*

This has been reiterated in the phase 2 review meetings with particular reference to SATs concerns. In relation to SATs results and school-based assessments, phase 1 and 2 teachers reported mixed outcomes, with few being enthusiastic about the programme's outcomes and others less certain.

What was being compared in the trial

Fidelity of implementation

IGR was implemented with varied fidelity across different schools and teachers. Most common variations observed or reported in the monitoring log included delivering IGR out of the classroom (e.g. often in the school library), confusing the teacher and TA roles (TAs in IGR have a strictly supportive role), the delivery of fewer sessions than the two teacher-and two TA-led per week, and not following the IGR lesson routine (by leaving out or adding steps or altering the sequence).

The programme team took action to restore such departures, and in mid-phase 1 a table with acceptable and non-acceptable variations was produced and shared with the teachers (appendix 4). An example of an acceptable variation was using one book per week (instead of two) for groups that required more time and spreading the routine across two sessions. On the other hand, delivering IGR in pull-out sessions was discouraged. This table was also used to inform the phase 2 programme support, with the only exception that delivering some IGR sessions during assembly was offered as a last organisational option to schools that felt keeping fully to the suggested organisation was not practically possible (with just 2 schools out of 15 using this option consistently in phase 2).

The IGR programme routine was also very particular. Each teacher-led IGR session was designed to have the same structure, including a discussion and game about the previous book, a new book introduction using story-telling techniques, a second game to familiarise pupils with the unknown vocabulary of the new book, the reading of the

new book with opportunities for collaborative problem solving, and finally a phonics game. The fine-tuned strategies involved in each step made IGR challenging to be implemented to high quality, and the programme team reported that only a handful of teachers mastered the approach fully (details in section 7). Common mistakes included: rushing or completely missing the introductory discussion; taking the element of surprise and memorisation out of the games; failing to engage pupils with storytelling; using inference questions; not giving enough time for individual problem solving or opportunities for collaborative problem solving; misusing or missing completely the phonics game; running out of time; not building links between the teacher and the TA session. Time issues were also reported across phases. For some teachers 30 minutes was not long enough for the IGR routine, and in such cases the phonics game was consistently left out.

The programme team reported better fidelity to the programme in phase 2 than in phase 1, but most problems were common across both phases. The quality of IGR teaching was reported to be improved between phases but not in a way that was consistent across all teachers.

The nature of the IGR approach and the number of support visits put pressure on teachers. It is indicative that the only negatively rated item in the phase 1 CMO questionnaire was the item: 'keeping on top of the programme is demanding for me' (appendix 9a). Most teachers explained in their interviews that their rating reflected various factors – the restrictions of the controlled trial design; keeping the IGR groups unchanged for the duration of the year; making sure that all the sessions were delivered with the teacher in the main role; and having to fill in a log. Yet, this could also suggest that teachers were not feeling fluent in using the programme and the strategies involved, perhaps revealing a training and practical experience issue. The same item was rated positively, but very low – almost neutral – in the phase 2 CMO questionnaire (appendix 9b). This could indicate that in phase 2 the revised training gave teachers more confidence compared to phase 1, but not enough to use the programme's strategies with fluency. This can be further related to the teachers' understanding of the theory behind the programme, discussed in the next section.

A limited understanding of the theory behind the programme

It is not clear to what extent the majority of the phase 1 and 2 teachers understood the theory behind the IGR programme (Appendix 1). This theory had been presented in the main and follow-up trainings and was reiterated in programme support visits and feedback to justify the teaching decisions and strategies involved.

There is some evidence that teachers engaged with aspects of the theory, for example phase 1 and 2 teachers noted that using the Lotto game enabled pupils to be familiar with the vocabulary of the text, before doing the actual reading and that it was a useful strategy that allowed pupils to be relaxed and engaged during the reading of the new text. Some teachers also used the actual IGR games or specially constructed ones with the rest of their class. However, there were teachers (especially in phase 1) who misinterpreted the role of the Lotto and used it after the reading of the new text to provide some context to the words (these links have already been made at this point). Other teachers also used the Lotto in an analytical way, i.e. to make grammatical and phonics comments, moving away from the simplicity central to IGR.

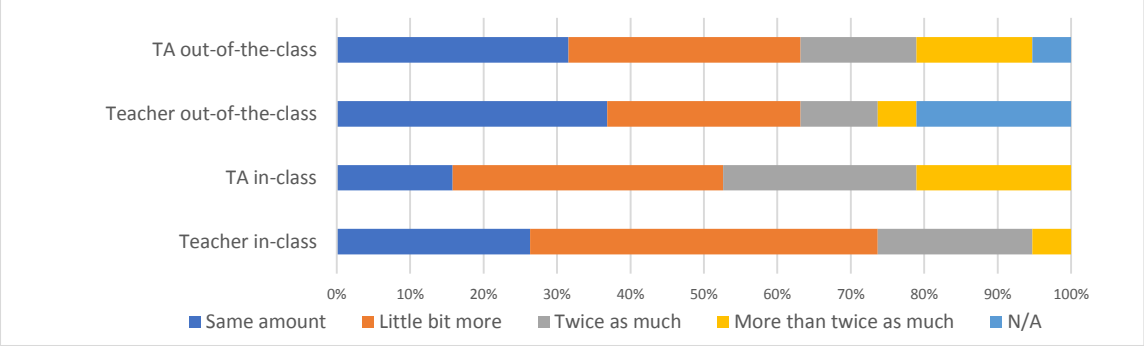
Yet, in the teacher interviews, it was made clear that the teachers mostly focused on finding solutions to the practical issues of the programme's delivery (staffing and organisational matters), and some thought of the programme as resources ready to use and less as a particular methodology based on research-evidenced practice. This could also be deduced from teachers' responses about their future plans, when most of them reported plans to use the IGR resources in a looser way, not just in terms of the organisation (which was inflexible due to the RCT design), but crucially not with reference to the programme strategies. This could suggest, on the one hand, that many teachers implemented the programme in a mechanical way, without engaging in the teaching implications, and, on the other, that training failed to enable teachers to understand the underpinning theory of the programme's methodology and value its importance. This is important because of the complexity associated with the programme teaching routine and strategies.

Control teaching

The control teaching data were collected from the phase 1 control schools, since there was no control group in phase 2. Of the 33 control teachers 29 responded to an online survey sent in November 2015. The majority of the respondents (28 out of 29) reported using a group reading organisation, with pupils organised according to similar reading levels (23 out of 29 cases). When the teacher was reading with one of the groups, the rest of the class was working on independent work (16 out of 29 cases) or with another teacher or TA or using some sort of combination. Programmes or approaches used in the schools included: *Reading Recovery*, *the "5 Minute Box" for literacy*, *ReadWriteInc*, *Shrewsbury 'Bookfest' materials*, *Purple Mash*, *Toe by Toe*, and *Reciprocal Teaching*.

A second survey was sent at the end of phase 1 (autumn 2016), with a more specific focus on the 4 identified control pupils. The first question was whether the project's identification of these pupils had any bearing on how the teachers taught them. Of 19 teachers 10 responded that although they remembered the pupils, this did not have any bearing on their teaching. The rest of the teachers (9) believed that their teaching had been affected to some extent (6 out of 19); or they had completely forgotten the identified pupils (3 out of 19). When teachers were asked how much time they themselves and their TAs were spending with the identified pupils, it was clear that a lot of additional teaching time was given to the identified pupils (Figure 3).

Figure 3. Control teaching: time spent with identified pupils compared to class pupils



When teachers were asked what kind of additional support they offered to pupils needing it, they reported: 1:1 reading opportunities with a TA (17 out of 19 teachers), extra phonics (16 out of 19), 1:1 reading opportunities with a teacher (12 out of 19), various differentiated activities (9 out 19), extra comprehension (8 out of 19), and other literacy programmes (6 out of 19).

The two teachers who discussed their responses further in a phone interview explained that the identified pupils in their classes had difficulties in reading and thus they had additional support. This involved in the one school more than twice as much time when it came to TA support, whereas in the other, all 4 identified pupils read every day with the teacher or a TA inside or outside of the regular class. Both schools offered extra phonics support to all or some of the identified pupils, with one using the "5 Minute Box" for literacy and the other Toe by Toe as a daily, one-to-one intervention, delivered by a TA early in the morning or out of the class (these are both synthetic phonics programmes).

13. Phase 2 teacher case study analysis

The main questions that will be explored in these case analyses are:

For cases where there was a match between level of IGR fidelity and reading gains

a. For high fidelity-high reading scores (H-H)

   Is IGR teaching and/or some other factors related to higher than mean reading gains?

b. For low fidelity low-low reading score (L-L)

   Were low reading gains because IGR was used poorly or some other factors that were related to lower reading gains?

For cases where there was a mismatch between IGR fidelity and reading gains

c. For high or medium fidelity-low reading scores (H/M-L)

   Why are the pupils not getting higher reading gains, even though the teacher was using IGR at an average/ high level?

d. For low or medium fidelity-high reading scores (L/M-H)

   What other factors despite low or medium fidelity teaching were related to higher reading gains?

Methods

The selection of cases was based on a combination of fidelity scores (4-point scale calculated from programme team observations using the fidelity index in appendix 7) and mean group reading gain scores from the SWRT (SWRT was seen as more independent than Hodder gains). This procedure could not be followed for phase 1, as there was less accurate phase 1 programme observation data.

A scatter plot was produced with fidelity scores (x axis) and SWRT mean group gains (y axis) for the phase 2 teachers who were interviewed for the process evaluation. Using the fidelity and reading scores, different combinations were selected to represent teachers along this continuum, and the red dots represent the selected teachers. Different combinations are presented in table 19, and detailed scores are given in table XX. The different combinations are: high-high, high-low, middle-high, middle-low, low-high, low-low.

Figure 4. Scatter plot: Phase 2 teachers' fidelity and mean group reading gains (SWRT)

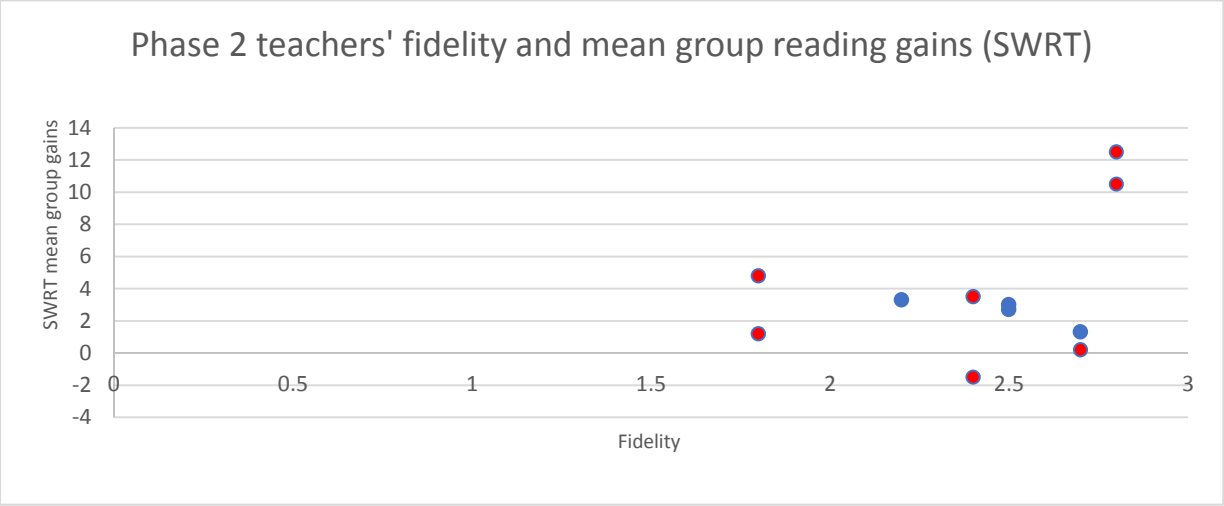


Table 19. Phase 2 teachers: combinations of fidelity and mean group reading gains (SWRT)

	Fidelity scores – mean 2.4			SWRT scores – IGR group 3.7	
Teacher name	High – **above mean**	Middle – **around mean**	Low – **below mean**	High – **above mean**	Low – **below mean**
T1	✔			✔	
T2	✔			✔	
T3	✔				✔
T4			✔		✔
T5			✔	✔	
T6		✔			✔
T7		✔			✔

Table 20. Phase 2 teachers: fidelity and mean group reading gains (SWRT)

(L=low; M=medium; H=high)

Teacher name	Fidelity	SWRT group mean difference	Interview	Observation
T1	2.8 H	10.5 H	yes	No
T2	2.8 H	12.5 H	yes	No
T3	2.7 H	0.2 L	yes	Yes
T4	1.8 L	1.2 L	yes	Yes
T5	1.8 L	4.8 H	yes	No
T6	2.4 M	3.5 L	yes	Yes
T7	2.4 M	-1.5 L	yes	Yes

Framework used in the case analyses

All 7 cases were analysed in terms of factors relevant to IGR group teaching and the group of pupils' reading score changes over the period of the two-term trial. The table below sets out these broad factors, what they cover and what data sources were used in the case analyses.

Table 21. Framework used in teacher cases

Factors	Data sources
1. Teacher characteristics: Education, years of experience, age, roles in school (e.g. SENCo), job-sharing, self-efficacy scores etc.	Demographic and self-efficacy questionnaires, information from programme team and independent visits
2. School characteristics: Area, FSM, literacy provision, school leaders and support, other information etc.	Publicly available demographic data, information from programme team and independent visits
3. IGR pupil characteristics: Year group, Pupil Premium, gender, EAL, SEN, reported issues (such as absences, behaviour etc.)	Demographic questionnaire, teacher interviews, the log
4. IGR organisation: IGR organisation, attitude towards IGR organisation	Independent observations, teacher interviews, the log
5. IGR teaching: Brief summary of IGR observations	Programme team and independent observations
Attitude towards IGR teaching, reported issues, pupils' engagement response to IGR	Programme team and independent observations, teacher interviews, pupil interviews, the log
6. The rest of the class: Using IGR with other pupils, concerns arising from IGR delivery (for class pupils)	Teacher interviews, the log
7. Mean IGR group reading scores, and teacher reported outcomes	SWRT and Hodder mean gains, reported outcomes from interviews and CMO
8. Summary: how analysis addresses relevant question for the case	

Teacher 1 (T1): High fidelity-High gains

Teacher characteristics

The teacher was a 31-year-old man with a Post Graduate Certificate in Education (PGCE) and 5 years of teaching experience. He reported having undertaken in-service training in early literacy teaching. His self-efficacy for teaching reading mean score in the questionnaire reported at the training day was 8.5/9 (the mean across all the teachers was 7.1/9) and remained stable throughout the year (8.3/9 at the end-of-the-year review meeting – the mean was 7.8/9).

School characteristics

He was serving in a suburban school in the West Midlands with 31.4% of pupils in receipt of Free School Meals, one of the highest percentages in the study. The school had strong literacy provision, and keen leaders who supported the programme – e.g. the Assistant Head (responsible for literacy) offered to read with one of the non-IGR reading groups (and did this for both participating classes), so that all pupils had equal reading time with a teacher. This happened every week, all year round, and he noted in his interview that: '*if I wasn't in this school and somewhere else and didn't have that available, I would have found it really difficult to fit in every child reading to me each week*'.

The local Literacy Adviser was very supportive of the programme and visited the school several times to make sure that the teachers were using the programme with good fidelity.

IGR pupil characteristics

He taught an IGR group of four Year 3 boys, of whom none had English as an additional language (EAL), one was identified as in receipt of Special Education Needs (SEN) School Support, and another had a Speech Language and Communication Needs (SLCN) Education Health and Care plan (EHCP). No pupil was eligible for Pupil Premium. The pupils' phonics scores from the statutory phonics test from the end of Year 1 ranged from 13 to 22, with 32 being the threshold.

Two of the pupils were reported to be often late, disorganised or away due to family circumstances.

IGR organisation

During Guided Reading, there were 5 groups plus the IGR group. The teacher was concerned about the noise (IGR pupils were particularly enthusiastic in playing the games), so on a rotating basis one of the groups was working just outside of the class with a TA – this was usual practice in the school. This was possible because two TAs were available: one was working with the group out of the class, and the other was overlooking the class during the teacher-led IGR sessions. With the added support of the Assistant Head, the IGR organisation was very smooth.

IGR teaching

The teacher was observed by the programme team to have good fidelity to the programme (2.8/3). There was no independent observation. He appreciated the multi-perspective approach of IGR and noted that it addresses all the different aspects of reading – however, he also felt that it took him some time to get used to the programme, and to the amount of detail involved in the programme delivery.

The rest of the class

The class had 29 pupils of whom 9 were girls, 9 eligible for Pupil Premium, 8 had EHCP plans, 2 had SEN statements and 8 were identified as in receipt of SEN School Support (this includes IGR pupils). There were no pupils with EAL. During the year no programme materials were used with class pupils, but the teacher noted that children took an interest in them.

Mean group reading scores, and reported outcomes:

For the 4 IGR pupils, the mean gain was 8 standard score points on the Hodder test, and 10.5 on the SWRT. The mean ratio gain for both tests was 2.4 often seen as 'useful impact' (table below).

IGR group					
HGRT standard t1	**HGRT standard t2**	**ratio gain**	**SWRT standard t1**	**SWRT standard t2**	**ratio gain**
82	94	2.2	85	104	3.4
87	100	3.8	88	103	3.1
93	90	0.7	90	98	2
89	99	3.1	92	92	0.7
HGRT ratio mean gain		**2.4**	**SWRT ratio mean gain**		**2.3**

The teacher also reported in his interview positive pupil outcomes in reading expression, accuracy, comprehension and confidence. He particularly noted how the safety of the IGR group helped the self-esteem of the particular pupil with the SLCN EHCP who was not worried about exposing his difficulties in the small group setting. These findings were further supported by the teacher's CMO questionnaire responses in which all programme outcomes were rated highly positively.

The teacher also highlighted that according to school-based standardised tests the IGR group improved a lot more than the rest of the class. This was supported by our findings, as below.

Whole class		
HGRT standard t1	**HGRT standard t2**	**ratio gain**
99.2	103.9	**1.8**

Summary:

Is IGR teaching and/or some other factors related to higher than mean reading gains?

In this case, the IGR organisation was smooth and stable throughout the year, and this was partly because of the support of the school leaders and the availability of two TAs, and partly because IGR delivery fitted well with existing arrangements in the school. In addition, the teacher was positive towards the programme, and managed to engage the four boys in the IGR group who were reported being enthusiastic about the programme activities. The local Literacy Adviser (and programme team) also made several visits to make sure that the programme was used with high fidelity.

Factors related to IGR implementation and outcome: Teacher and pupil enthusiasm, school leaders actively engaged, Local Adviser support, school ethos, fitting well with existing guided reading arrangements

Teacher 2 (T2): High fidelity-High gains

Teacher characteristics

The teacher was a young female teacher with a Post Graduate Certificate in Education (PGCE). She did not provide precise information about her age and teaching experience and had not undertaken additional training in early literacy teaching. Her teaching reading self-efficacy improved notably from 5.3/9 at the training day (the mean was 7.1/9) to 7.6/9 at the end-of-the-year review meeting (mean 7.8/9).

School characteristics

She was serving in a Greater London school with very high FSM percentage (43.6%). The school had strong literacy provision and dedicated school leaders. More particularly, the Deputy Head was actively involved in the programme procedures, consistently supported the participating teachers, and was interested in exploring ways of using the programme after the completion of the project. The Head Teacher was reported by another participating teacher to prefer that literacy and mathematics interventions are offered in the regular classroom and delivered by the class teacher (as opposed to TA-led withdrawal sessions). The school was continuing with daily ReadWriteInc sessions (just before the end of the school day), with arrangements not affecting IGR timetabling.

IGR pupil characteristics

Her Year 2 IGR group had three boys and one girl, of whom 3 had English as an additional language (EAL), one pupil was identified as in receipt of Special Education Needs (SEN) School Support, and one was eligible for Pupil Premium. The pupils' phonics scores in the phonics test from the end of Year 1 ranged from 21 to 36, with 32 being the threshold.

IGR organisation

There were four other reading groups in addition to the IGR group, doing comprehension or spelling activities, reading one-to-one with a TA or reading for pleasure.

She was particularly concerned about the time devoted to IGR teaching (especially closer to the SATs) and was worried that other pupils were missing opportunities to read individually with the teacher and TA. She had explored other options, e.g. reading with the IGR group in assembly time for one of the sessions but found such options difficult to implement consistently. This concern was also reported by the Deputy Head.

The teacher also felt initially that 30 minutes was too long for the rest of her Year 2 class to work meaningfully, and that non-IGR pupils were getting distracted and interrupting the teacher. However, these issues improved during the year.

IGR teaching

The teacher was reported by the programme team to have good fidelity to the programme (2.8/3), but there was no independent lesson observation. She reported in her interview that she felt that IGR supports the enjoyment of reading and used creative ways and games to support pupil learning.

Yet, she was worried that the programme focuses more on accuracy rather than comprehension which is particularly important for the SATs tests. Despite these concerns, the teacher could understand the reason why IGR had to stay simple, and why and how IGR differs from a usual Guided Reading session. She tried then to practice comprehension and especially inference out of IGR, in English sessions. Her approach ('*I don't want to turn IGR into a Guided Reading session'*) suggests that she was reflective in her teaching, and that the rationale and theory of IGR were discussed in the school (there were 4 participating teachers, and the actively involved Deputy Head).

The rest of the class

She was teaching a class of 31 children, of whom most were EAL pupils (yet with some exposure to English at home), 4 were identified for SEN School Support and 2 were eligible for Pupil Premium.

She reported that 'other children can become distracted at times as they are keen to see the games the children in the IGR group are playing. They have expressed that they would also like to use the resources in Guided Reading'. The teacher did try to use IGR resources with other class pupils, but felt they were not relevant to them (too easy, and not easily used with larger groups).

Mean group reading scores, and reported outcomes:

For the 4 IGR pupils, the mean gain was 6 standard score points on the Hodder test, and 12.5 on the SWRT. The mean ratio gain (table below) for both tests was 2.1 often seen as 'useful impact' (Brooks, 2016).

It should be noted that the pupils scored high on the Hodder test (above 94), but lower on SWRT (two of these pupils scored 76 and 85). SWRT was conducted independently by visiting researcher.

IGR group					
HGRT standard t1	**HGRT standard t2**	**ratio gain**	**SWRT standard t1**	**SWRT standard t2**	**ratio gain**
96	102	2.4	76	100	3.8
100	101	1.4	85	96	2.2
94	101	1.4	98	109	2.8
96	106	2.8	102	106	2
HGRT ratio mean gain		**2**	**SWRT ratio mean gain**		**2.2**

The rest of the class had modest mean ratio gains, but slightly less progress than the IGR group.

Whole class		
HGRT standard t1	**HGRT standard t2**	**ratio gain**
102.9	105.1	**1.7**

The teacher reported positive outcomes for the IGR group, especially in relation to their confidence and oral language skills. She also noted that they were working towards meeting the expectations for the end of KS1 but were still behind their classmates – this is confirmed by our findings as well (IGR mean scores from HGRT time 2: 102.5; rest of the class from HGRT time 2: 105.1). The teacher was only worried about the progress of a particular child who had difficult circumstances at home. She rated highly all outcomes in her CMO questionnaire at the end-of-the-year review meeting.

Summary:

Is IGR teaching and/or some other factors related to higher than mean reading gains?

This teacher demonstrated an understanding of the theory and rationale behind the programme, and she used IGR with fidelity despite the pressures of the SATs. This suggests, on the one hand, that she did not follow the programme in a mechanical way, and on the other hand, that the ethos of her school enabled and supported her to reflect on her teaching.

She was, however, particularly worried about the amount of time spent with the IGR group, and was not always comfortable with the IGR organisation. Yet, whilst nonetheless managing to maintain close fidelity to the IGR programme, she also offered a lot of additional support to her IGR pupils (especially in the final term), either as SAT's preparation (comprehension tasks and inference) or as phonics practice (RWInc programme) since some had failed their Year 1 statutory phonics screening.

Factors related to IGR implementation and outcomes: school leaders actively engaged, understanding of the theory behind the programme, additional input (phonics/ comprehension), and school ethos

Teacher 3 (T3): High fidelity-Low gains

Teacher characteristics

The teacher was a 39-year-old woman with a Post Graduate Certificate in Education (PGCE) and 5 years of teaching experience. She reported having undertaken in-service training in the area of early literacy. Her teaching of reading self-efficacy mean score on the training day was 8/9 (mean 7.1/9) and remained at a similar level 8.4/9 at the review meeting at the end of the year (mean 7.8/9).

School characteristics

She was serving in a school in a South-West Local Authority. The FSM percentage of the school was 26.3%. The school had a particular Guided Reading protocol that the teachers were following, articulating the minimum expectations, principles and values of reading in the school. The school also used the Accelerated Reader (AR) software programme (IGR pupils also used AR). Throughout the programme implementation, the project team communicated directly with the teachers, and there were minimal communications with the school leaders – though the two participating teachers reported that leaders were supportive of the programme.

IGR pupil characteristics

There were 2 boys and 2 girls in this Year 3 IGR group, and all 4 were eligible for Pupil Premium. Scores from the phonics test from the end of Year 1 ranged from 19 to 28 correct words read out, with 32 being the threshold. No pupil was identified for SEN support. The teacher reported occasional absences for some of the pupils.

IGR organisation

She had one TA available for her class, and during teacher-led IGR, she used to send one group out of the class (area still visible from the class) with the TA. In her interview, she explained that this was common practice in the school, and she found it difficult to have two adult voices in the class. She also noted that she was able to have good control

of the remaining groups in the class when she was delivering IGR – the independent observation confirmed this, as her classroom was one of the most settled classrooms with few distractions from non-IGR pupils.

The teacher had wondered whether this organisation would be acceptable for the evaluation and had sought guidance when she had started using the programme. The project team advised her to feel free to explore the different options for organising her classroom. In her interview, the teacher wondered whether IGR could work better in TA-led pull-out sessions, as pupils would be able to work without other distractions; however, she did not use this organisation during the trial.

IGR teaching

The teacher was observed to have generally good programme fidelity by the programme team (2.7/3) with some notable weaknesses especially in the collaborative reading element. The independent observation indicated that the teacher was trying to be faithful to the programme, but used the programme strategies in a mechanical way. Her approach was also reflected in her body language: she was facing directly the two weaker pupils (to support them as far as possible), and was sitting far away from the two stronger members of the group who remained unengaged. This worked against developing the good group dynamics that are central to IGR. This can also be related to her way of using storytelling that was observed to be closer to an inference-based questioning (Guided Reading protocol), failing to engage pupils in the story of the new book.

The teacher reported that she felt IGR supported pupils' vocabulary and language, but was worried that there was not enough focus on comprehension to prepare the pupils for the SATs test. She was then planning to include more opportunities for comprehension practice closer to the summer term.

The rest of the class

There were 29 class pupils, of whom 19 were boys, 14 were eligible for Pupil Premium, 7 had English as an additional language (EAL), 2 had an EHCP, and 2 a statement of SEN. They were organised into 5 reading groups (including the IGR group). The teacher was using IGR resources with a non-IGR group, with an emphasis on inference and deduction as required by the National Curriculum. Towards the end of the year, she reported in the log that IGR was used as a SEN/ EAL intervention.

Mean group reading scores, and reported outcomes:

For the four IGR pupils, the mean decrease of 8.8 standard score points on the Hodder, and a slight gain of 0.2 on the SWRT (table below), suggest that IGR had little or no positive impact.

IGR group					
HGRT standard t1	**HGRT standard t2**	**Ratio gains**	**SWRT standard t1**	**SWRT standard t2**	**ratio gains**
73	73	0.2	70	81	1.7
82	82	0.8	83	74	0
99	85	-1.4	97	98	1.4
104	83	-2.2	92	90	0.7
HGRT ratio mean gain		**-0.6**	**SWRT ratio mean gain**		**0.9**

The rest of the class also did not record any gains on the Hodder test, as below.

Whole class		
HGRT standard t1	**HGRT standard t2**	**ratio gains**
100.2	100.2	**1.1**

However, the teacher reported positive outcomes for the IGR group, especially in relation to their confidence, but she noted that '*most of them are confident [because they are] in a small group and can make mistakes*' without having to worry about other pupils' judging their reading. She reported high programme outcomes in her CMO questionnaire.

Summary:

Why are the pupils not getting higher reading gains, even though the teacher was using IGR at an average/ high level?

She implemented IGR with a rather mechanical approach. She tried to make sure that she followed the study's protocol in relation to the programme routine and organisation, and she checked with the programme team for the adjustments that she made (e.g. in relation to her class organisation).

However, she was observed to be less able to use the programme strategies effectively to engage her IGR group at a group level using the different reading levels of the pupils as a teaching tool – i.e. supporting collaborative problem-solving skills, modelling of reading, and boosting the confidence of the more able readers. As a result, the pupils were observed to be less enthusiastic, with some still struggling and others being bored by the activities.

This points also to limitations about the way teaching fidelity was measured, as the teacher followed the procedures and routines (and thus checked many fidelity boxes), but she did not achieve engagement with the strategies involved to some depth. E.g. the teacher scored lower on the critical element of storytelling but, since all items were weighted equally, she received a high overall score.

Factors related to IGR implementation and outcomes: mechanical approach, pupil engagement at an individual but less at a group level, less faith in the theory behind IGR and classroom organisation, SATs pressures, school leaders not directly involved

Teacher 4 (T4): Low fidelity-Low gains

Teacher characteristics

The teacher was a 47-year-old woman with a Post Graduate Certificate in Education (PGCE) and 5 years of teaching experience. She reported having undertaken training in the area of early literacy. Her self-efficacy mean score on the training day was 6.7/9 (mean 7.1/9) and improved (mean 8/9) by the end of the year (mean across teachers 7.8/9).

School characteristics

She was serving in a South-West school with 26.3% FSM percentage (same school as T3). The school used a Guided Reading protocol articulating the minimum expectations, principles and values of reading in the school. During the programme implementation, the project team communicated directly with the teachers, and there were minimal communications with the school leaders who were reported by the teachers to be supportive to the programme.

IGR pupil characteristics

This Year 2 IGR group had 2 boys and 2 girls, of whom all 4 pupils were eligible for Pupil Premium, and 1 was identified for SEN School Support. Their scores from the phonics screening from the end of Year 1 ranged from 32 to 36, with 32 being the threshold.

IGR organisation

Pupils were organised into 5 reading groups (including the IGR group). The teacher reported in her log that 2 of the groups worked with TAs (two TAs were often available) and 2 on independent work. T4 noted that Year 2 children needed to be trained to work well independently for 30 minutes, but this was not a matter of concern. She was only worried that '*the practicalities of the programme are tricky in that it's four sessions with two adults, so for me to get round the other groups of my class takes me two weeks as opposed to one week*'. Despite her concerns about the teacher time spent with the IGR group, she kept to the suggested organisation for the duration of the trial.

IGR teaching

The teacher was observed by the programme team to have poor fidelity to the programme (1.8/3), particularly in relation to the storytelling aspect. During the independent observation, the teacher read part of the book instead of narrating it and failed to deeply engage the pupils in the new story; pupils seemed to enjoy the session but were not enthusiastic. The programme team reported that T4 gradually improved and managed to use the approach effectively towards the end of the year.

The teacher believed that what was useful in IGR was '*the structure and the pace. It means that we get through a lot of books and material and there is quicker progress because of that'*.

She also compared IGR to usual Guided Reading: '*I think the useful side of it is it's really focused me on actual reading and phonic skills as opposed to Guided Reading as a whole where we discuss features of books. So, it's very much about them acquiring language and it's focused me on SATs*'.

The rest of the class

The teacher taught a class of 29 pupils, of whom 15 were girls, 14 eligible for Pupil Premium, 3 EAL and 3 were identified for SEN School Support (including the IGR group pupils).

The teacher did not use IGR materials with other groups, mainly because of time limitations.

Mean group reading scores, and reported outcomes:

For the IGR group, the mean gain was 16 standardised points on the Hodder test, and 1.2 on SWRT (table below). The difference between the tests is significant; the high HGRT gains seem to be due to the low baseline scores (as low as 72), suggesting possible difficulties in using HGRT at time 1, whereas the SWRT scores show little progress (a ratio gain of 1.2). It should be noted that HGRT was delivered by the class teacher, whereas SWRT by a trained visiting Research Associate.

IGR group					
HGRT standard t1	**HGRT standard t2**	**ratio gains**	**SWRT standard t1**	**SWRT standard t2**	**ratio gains**
76	113	4.5	102	101	1.2
102	104	2.2	105	105	1.4
91	106	3.4	92	99	1.4
98	108	2.8	107	106	0.8
HGRT ratio mean gain		**3.2**	**SWRT ratio mean gain**		**1.2**

The class pupils improved on the HGRT – ratio gain was 1.8 (often described as 'modest' gain).

Whole class		
HGRT standard t1	**HGRT standard t2**	**ratio gain**
101.5	110.28	**1.8**

She reported in her interview that pupils in the IGR group were progressing well, and rated highly most programme outcomes in the CMO questionnaire (a mean of 1.5 on a scale ranging from -2/+2).

Summary:

Were low reading gains because IGR was used poorly or some other factors that were related to lower reading gains?

The teacher was observed to have poor fidelity to the programme. The programme team reported improved fidelity, but this only happened towards the end of the year.

It is not clear whether the teacher had a good understanding of the underlying principles of the programme. She noted that IGR helped her focus on phonics and SATs preparation (as opposed to reading comprehension that she related to usual Guided Reading sessions), but the aim of IGR is to re-engage struggling pupils with reading, taking away the anxiety of past failures. Thus, IGR should not be seen as an assessment preparation method, but as a sophisticated teaching approach for pupils who could not be helped through other avenues. This can indicate that the teacher misinterpreted the goals of the programme, which might partly explain the implementation difficulties she faced.

The mismatch between Hodder and SWRT might be due to test implementation issues at baseline, suggesting that weak IGR implementation can be related to poor reading gains.

Factors related to IGR implementation and outcomes: Unclear understanding of the theory behind IGR, concerns about IGR organisational model, poor programme implementation for most of the year, school leaders not directly involved

Teacher 5 (T5): Low fidelity-Medium high gains

Teacher characteristics

The teacher was a 53-year-old woman holding a BEd with QTS and had 10 years of teaching experience. She reported having undertaken in-service training in the area of early literacy. Her self-efficacy mean score on the training day was 6.6/9 (mean 7.1/9) and improved throughout the year (7.9/9 at the end-of-the-year review meeting – mean 7.8/9).

School characteristics

The teacher was serving in a South-West school with 8.8% Free School Meals percentage. The school run a School Direct Teaching Programme supported by a local University, which also meant that trainee teachers were assisting in the participating classrooms. The school offered additional phonics and spelling TA-led pull-out sessions every morning and afternoon to weak readers (10min sessions), including some of the IGR pupils. Throughout the year, Friday was dedicated to comprehension and all pupils had activities tailored to the reading they did during the week.

IGR pupil characteristics

The teacher taught a Year 2 group of 3 girls and 1 boy, of whom 2 were eligible for Pupil Premium, and 3 were identified for SEN School Support. Their scores on the phonics screening from the end of Year 1 ranged from 35 to 38, with 32 being the threshold.

IGR organisation

The teacher had 6 reading groups in her classroom (including the IGR group), but since she had a TA and a trainee teacher available to work with the rest of the class, she did not experience any issues. This is from her log entries: '*IGR now well established in classroom, no concerns; my only concern is the length of time non-IGR groups have to work independently*'. As she was teaching a Year 2 group, 30 minutes were often too long a period for class pupils to work independently.

She particularly appreciated the IGR organisational model, as she felt that it gave structure to the Guided Reading organisation to the benefit of all pupils:

'*We've got a sort of system up and running […] So for example we're doing adjectives this week, so they'll look through their Guided Reading book and be adjectives detectives and find them and list them and write them in sentences in context. So, the longer we're doing it, then the better it is, and the more it's benefiting the whole class rather than just the IGR four*'.

IGR teaching

The teacher was observed by the programme team to have poor fidelity to the programme (1.8/3), and there was no independent observation. What the programme team noted though was that the low fidelity scoring was because aspects of the routine were missing due to adjustments required to slow down the pace of the lesson to be more relevant to the pupils' needs. As a result, for example, there was no time for the phonics game at the end of the session. However, the IGR elements that were present were taught with high fidelity and attention to detail. The programme team noted that some of the changes were appropriate for the group, but this was not captured by the fidelity index.

The pace of the lesson was an issue for the teacher who noted in her log that: '*my concern is, it does seem to be a little rushed to cover 2 books a week*'. She tried then to use only one book per week, but it took her some time to feel comfortable with the programme routine for one book. He teacher appreciated the IGR approach that allowed her to focus more on pupils' reading:

'*[In previous years] I used mainly my own material and sentence work. The children weren't always reading as such, they didn't have the book there all the time. Sentence work relating to whatever we were doing. So, there wasn't that focus on Guided Reading without IGR*'.

The rest of the class

The teacher taught a class of 30 pupils, of whom 17 were girls, 5 pupils eligible for Pupil Premium, and 5 EAL (including the IGR group pupils). No pupils were identified for SEN support. One of the non-IGR groups (a low-ability group) was using IGR resources with the support of a TA.

Mean group reading scores, and reported outcomes:

For the IGR group, the mean gain was 12.3 standard score points on the Hodder test, and 4.8 on SWRT (table below). The children in the group progressed at different rates (ratio gains ranging from 0.7 to 2.5 on the SWRT), and especially one child progressed ahead of the others affecting the mean. This group scored higher compared to other groups in the study (scores in baseline as high as 114).

IGR group					
HGRT standard t1	**HGRT standard t2**	**ratio gain**	**SWRT standard t1**	**SWRT standard t2**	**ratio gain**
91	98	2.2	90	90	0.7
95	109	3.1	99	101	1.4
91	111	**4.1**	104	117	**2.5**
100	108	2.4	114	118	1.7
HGRT ratio mean gain		**2.9**	**SWRT ratio mean gain**		**1.5**

Class pupils improved on the HGRT with a ratio gain of 1.7 (often described as 'modest' gains).

Whole class		
HGRT standard t1	**HGRT standard t2**	**ratio gain**
110.3	118.2	**1.7**

The teacher rated highly most programme outcomes in the CMO questionnaire (a mean of 1.8 on a scale ranging from -2 to +2). She also reported positive outcomes in her interview, as illustrated by the excerpt below describing the progress of an IGR pupil:

The Year 2 children have a buddy partner with Reception children […] and the Year 2 will read a story to the reception child. [An IGR pupil] was reading this story to the four-year-old so beautifully, and we looked at each other and thought: well, she must know the story really well.

Summary:

What other factors despite low or medium fidelity IGR teaching were related to higher reading gains?

The teacher tried to slow down the pace of the lesson, introducing one book a week and this created some difficulties in using certain aspects of the routine (such as the phonics game) – difficulties that were captured by the fidelity index and resulted in lower scores.

The class organisation, however, was smooth as a trainee teacher was available for most of the year (in addition to the class TA) to assist with the rest of the class during teacher-led IGR. The teacher appreciated the organisational model of IGR that she felt gave a tighter structure to the way reading was organised and taught to all pupils in her class.

IGR pupils had, finally, a lot of daily and weekly additional input (e.g. phonics and comprehension) out of IGR that reflects also the ethos of the school. IGR pupils though progressed at different rates, with one child progressing ahead of the others.

Factors related to IGR implementation and outcomes: Particular school ethos, additional input for IGR pupils, appreciation of IGR organisation model, having a trainee teacher to assist with the rest of the class, IGR group with high baseline scores

Teacher 6 (T6): Medium fidelity-Low gains

Teacher characteristics

The teacher was a 56-year-old woman with 34 years of experience, the most experienced teacher in the study. She held a Bachelor of Education and reported having undertaken early literacy training. Her self-efficacy mean score in the questionnaire completed at the training was 8.9/9 (mean 7.1/9) and remained stable during the year (8.9/9 at the end-of-the-year review meeting – mean 7.8/9).

School characteristics

The teacher was serving in a West Midlands School with 14.2% percentage of Free School Meals. There were no direct communications between the school leaders and the project team.

IGR pupil characteristics

The teacher taught a Year 2 group of 2 girls and 2 boys, of whom 2 were eligible for Pupil Premium, and 1 was identified for SEN School Support. Their scores in the phonics screening from the end of Year 1 ranged from 32 to 37, with 32 being the threshold.

They all had interventions for oral language support (e.g. *Black Sheep*), vocabulary (e.g. *Word Aware*) in addition to IGR and one of the children had also TA-led phonics sessions.

IGR organisation

The IGR group read 1 book per week, and the teacher split the IGR routine using the 1st teacher-led session to play GoFish and introduce the new book, and the 2nd teacher-led session to play the Lotto, read the new book and play SWAP. This is not the suggested organisation for 1 book per week (half book in each session) and thus reduced the number of TA-led sessions: the IGR group had only one TA session every Friday inside the classroom but during assembly (the other pupils were away). This was partly because the new book was read in the 2nd session, so there was no material to be covered by the TA after the 1st session, and partly to give time to the teacher and TA to read with all other groups. This is a matter that was not captured well by the fidelity index which was focused on the teacher sessions. The teacher was asked to restore the number of TA sessions (2 in total), but argued that '*it would mean the rest of the class are not heard; it would have a knock-on effect*'.

The teacher also noted in her log that without an experienced TA it is difficult to implement the programme properly: *'When I'm doing Guided Reading I'm still looking around the class, I'm still having to manage the rest of the class. […] My TA is still young and she's not that competent yet to managing another 24 children'.* She also added that when the TA was away (to cover other classes or due to illness), it was not practically possible to use the IGR organisation.

IGR teaching

The teacher was observed by the programme team to have moderate fidelity to the programme (2.4/3) – which represents the mean fidelity score across teachers. The independent observation confirmed that not all aspects of the routine were present (resulting in lower scoring) because the teacher used to split her routine for 1 book per week without following the suggested procedures. The aspects of the routine that were present were mostly used with good fidelity, with some weaknesses observed in the storytelling and collaborative reading element.

The teacher appreciated the multi-perspective approach of IGR to reading, particularly the games and the phonics element of the programme: '*Last year we tended to just use books, so it's nice to actually have the different games and the phonics […] They're getting much more confident at recognising those sounds*'. Yet, she noted that future versions of IGR could include synthetic phonics and made-up words for pupils who have failed their Year 1 phonics screening.

Towards the end of the year (close to the SATs), she wrote in her log: *'I am worried they are not having comprehension written activities in preparation for SATs, as the week is taken up doing IGR activities, so no time to work with them on comprehension written skills'.* She also added in her next and final log that: *'This could affect their performance in SATs'.*

The rest of the class

The teacher taught a class of 29 pupils, of whom 13 were girls, 5 pupils eligible for Pupil Premium, and 9 EAL (including the IGR group). 4 pupils were identified for SEN School Support. Pupils were organised into 5 reading groups (including the IGR group). The teacher used IGR materials with one EAL boy with slow reading progress. Her main concern in relation to the class pupils was the limited time to read with them: '*Just time to focus on their reading, as they have to work independently and [there is] not always time to hear them'*

Mean group reading scores, and reported outcomes:

For the IGR group, the mean gain was 7 standard score points on the Hodder test, and 3.5 on SWRT (table below). The ratio gains suggest 'modest' impact of IGR.

IGR group					
HGRT standard t1	**HGRT standard t2**	**ratio gain**	**SWRT standard t1**	**SWRT standard t2**	**ratio gain**
101	94	0.1	92	90	0.8
96	110	3.2	90	94	1.5
99	116	2.1	92	104	1.8
89	93	2.1	101	101	1.2
HGRT ratio mean gain		**1.8**	**SWRT ratio mean gain**		**1.3**

The same applies to the class pupils.

Whole class		
HGRT standard t1	**HGRT standard t2**	**ratio gain**
105.5	108.2	**1.7**

The teacher rated highly most programme outcomes in her CMO questionnaire (a mean of 1.5 on a scale ranging from -2 to +2). She also reported positive outcomes in her interview. In relation to the pupils' progress she noted that: '*The confidence, the fluency has really improved, the discussion about characters, what's happening, all that sort of book knowledge and language has really developed*'. She also described the IGR group as a nurture group for pupils with low confidence: '*It's such a small little nurture group, I think it's really developed those you know lacking in confidence to become more confident in saying something*'.

Summary:

Why are the pupils not getting higher reading gains, even though the teacher was using IGR at an average/ high level?

In her attempt to use one book per week, the teacher changed the programme routine in a way that also affected the number of her TA sessions, thus creating a dosage problem. This matter was not reflected in her fidelity scoring, as the fidelity index did not take into account TA-related matters. The teacher also reported difficulties in relation to her TA who was young and less experienced and could not assist effectively during teacher-led IGR.

There is evidence that the teacher had a good understanding of the theory and rationale behind the programme (e.g. she saw the IGR group as a nurture group), but she was also particularly concerned about the SATs requirements. This is evident from her log entries, but also from the interventions that the IGR pupils had in addition to IGR. It is not clear how her teaching was affected by the SATs, as there are no observations from this particular period.

Factors related to IGR implementation and outcomes: difficulties with IGR organisation, TA-related issues, dosage issues (half TA consolidation time), difficulties with using the full IGR routine, SATs pressures

Teacher 7 (T7): Medium fidelity-Low gains

Teacher characteristics

The teacher was a 38-year-old woman with a Post Graduate Certificate in Education (PGCE) and 12 years of teaching experience. She reported not having undertaken in-service literacy training. Her self-efficacy mean score on the training day was 7.6/9 (with the mean across teachers being 7.1/9); the teacher did not attend the end-of-year review meeting and did not complete a 2nd self-efficacy questionnaire. She job-shared her class and did not complete any fortnightly logs.

School characteristics

She served in a Greater London school with high FSM percentage (43.6%) (same school as with T2). The school had strong literacy provision and dedicated school leaders. E.g. the Deputy Head was actively involved in the programme procedures, consistently supported the participating teachers, and was interested in exploring ways of using the programme after the completion of the project. The Head Teacher was reported to prefer for literacy and other

interventions to be offered inside the regular class and delivered by the class teacher (as opposed to TA-led withdrawal sessions).

IGR pupil characteristics

The teacher taught a Year 3 group of 2 girls and 2 boys, of whom all pupils were identified for SEN School Support and none was eligible for Pupil Premium. 3 of the pupils were EAL. Their scores from the Year 1 phonics test ranged from 24 to 29, with 32 being the threshold.

IGR organisation

Pupils were organised into 6 reading groups (including the IGR group). The teacher reported in her interview that in order to have an IGR group of 4 pupils, the other groups had to be larger. This was because this school had particularly large classes (35 pupils), compared to other schools in the study. Having larger groups allowed the teacher to read with all pupils every week – yet, she still had to read with one of the (non-IGR) groups in assembly time:

'*We've made our groups bigger, it means we get through them all. There's one group that I have to read with during an assembly time Monday assembly time*'.

IGR teaching

The teacher was observed by the programme team to have average fidelity (2.4/3) with some weaknesses in the storytelling and collaborative reading aspects. The independent observation noted particularly the calm manner of the teacher, and the enthusiasm of the group. There were also interesting dynamics in the classroom, with the calm nature of the teacher being complemented by a particularly lively teaching assistant who took over the rest of the class during teacher-led IGR. Yet, the programme team noted that these dynamics did not work as effectively with the job-sharing teacher who was also lively, and there were some personality clashes.

The teacher was positive about the IGR approach to reading and noted that it was particularly suitable for EAL pupils: '*the books are [more accessible to the children] than a lot of the books that we've got in school. The games repeating the language […] are really good for the EAL children'*.

The rest of the class

The teacher taught a class of 34 pupils, of whom 18 were girls, 6 pupils eligible for Pupil Premium, and 32 EAL (including the IGR group). 4 pupils had an Education Health & Care Plan (EHCP) for SEN. The other groups were having activities influenced by Reciprocal Teaching principles.

The teacher also occasionally used IGR materials with another group of 8 pupils:

'*I have done it a few times with [a larger] group and they are absolutely fine sharing the books. I've done the Lotto game with them, and they've shared one card between two, and I've done the Go Fish game where they've worked in partners to do it. So, it does work with a group of eight*'.

Her main concern in relation to the impact of IGR on the class pupils was that there was limited TA time to read with them, as almost all of them were EAL and could benefit from one-to-one support:

'*On the two days that the [IGR] children read with the TA that means she's not getting chance to hear* [other] *individual readers. Because we've got so many [EAL] children, it really benefits the one-to-one readers. So that's a little bit difficult, they're not getting that one-to-one reading time'*.

Mean group reading scores, and reported outcomes:

For the IGR group, there was a mean decrease of -1.2 standard score points on the Hodder test, and -1.5 on SWRT (as below), suggesting that the programme did not have any impact on the IGR pupils.

IGR group					
HGRT standard t1	**HGRT standard t2**	**ratio gain**	**SWRT standard t1**	**SWRT standard t2**	**ratio gain**
102	99	0.2	98	103	1.7
82	80	0.5	83	78	0.4
77	84	0.7	85	82	0.5
96	89	0	89	86	0.8
HGRT ratio mean gain		**0.3**	**SWRT ratio mean gain**		**0.8**

Similarly, the Hodder test did not capture any progress for the class pupils.

Whole class		
HGRT standard t1	**HGRT standard t2**	**ratio gain**
99.4	96.4	**0**

The teacher did not complete a CMO questionnaire. In her interview, she reported that the IGR pupils became more confident because of the structure of the IGR approach: '*Children who are less confident really appreciate that they feel like they've had a heads up on it*'.

Summary:

Why are the pupils not getting higher reading gains, even though the teacher was using IGR at an average/ high level?

The teacher used IGR averagely and her job-sharing teacher was observed by the programme team to have good fidelity. Yet, it appears that the teachers had different teaching styles that were reflected also in the way they organised their cooperation with their class TA during teacher-led IGR.

The independent observation of the teacher found that the TA took over the class during IGR (whereas in other lesson observations TAs mostly worked with one group) which worked well with the calm manner of the teacher, but as reported by the programme team not with the personality of the job-sharing teacher. This suggests that the classroom dynamics during IGR might not be stable and settled throughout the year. This raises further questions as to which is the best approach to support a class in teacher-led IGR and what should be the TA's role – a replacement for the teacher or a person with a strictly assisting role.

The teacher also experienced difficulties in using the IGR organisational model in a large classroom of 34 pupils where groups often had 5-6 pupils. In order then to have a group of four pupils for IGR, the other groups had to become larger (some with 8 pupils), with implications for teaching.

Factors related to IGR implementation and outcomes: job-sharing teachers, teacher and TA roles, large class, and IGR-related organisation difficulties

Summary

Seven teachers with different profiles were discussed as individual cases representing different combinations of fidelity scores and reading gains. The above analysis suggests the following:

T1 and T2: high gains can be attributed to the quality of IGR teaching and other supportive factors

T3: low/no gains can be attributed to specific aspects of IGR teaching – this case also indicates that the fidelity index might not give due weight to the quality of collaborative reading activity

T4: the low gains can be attributed to the low level of IGR teaching and organisational issues

T5: the good gains can partly be attributed to other programmes operating during IGR, and partly to IGR teaching, as the fidelity index could not capture appropriate changes in pace. The IGR pupils had also higher initial reading scores compared to other groups in the study

T6: low gains can partly be attributed to implementation, class organisation and TA-related issues and partly to not using the full sequence and number of teacher/TA-led sessions

T7: low gains were not only to do with medium IGR fidelity (not high level) but large class, job-share and inconsistent IGR approach between job-sharing teachers

These cases are now discussed together to explore the relationship between teaching fidelity and reading gains and discuss the way fidelity was calculated in the study.

Match between level of IGR fidelity and reading gains (T1, T2, T4)

These teachers' case studies represent a relationship between high fidelity and high reading gains (T1 and T2) and low fidelity and low reading gains (T4) which is consistent with the original hypothesis that using the IGR programme with good fidelity corresponds with higher reading gains for the IGR group.

The main factors associated with high fidelity and high gains (from T1 and T2) were the enthusiasm of teachers and pupils, the involvement of the school leaders and local Literacy Adviser, the ethos of the school (as reflected in kind of leadership and approach to literacy provision), the way the IGR model fitted with existing reading organisational arrangements (Guided Reading or other model), the understanding of the theory and rationale behind the programme and the presence of additional to IGR input (in the form of phonics, comprehension or other) for the IGR pupils.

The main factors associated with low fidelity and low gains (from T4) were an unclear understanding of the theory and goals of the programme, concerns about its organisational model, difficulties in implementing aspects of the programme (e.g. storytelling) thus failing to engage and inspire pupils and serving in a school with distant school leaders.

These factors seem to suggest that having a good understanding about the theory, rationale and goals of the programme (from the teacher's perspective) and having the support and active engagement of the school leaders (from the school's perspective) seem to be central to achieving good results in both programme fidelity and reading gains. More particularly, the two teachers that had high fidelity and high scores both showed a good understanding about the theory behind the programme with the one teacher (T1) appreciating the multi-perspective approach of IGR to reading and the other (T2) emphasising the simplicity of the IGR approach and the differences with typical Guided Reading. Both teachers also served in very supportive schools, and their school leaders were actively involved in the programme procedures, supporting the teachers, helping with the delivery arrangements, organising for assessments and programme-related visits etc. On the other hand, T4 (Low fidelity-Low gains) was serving in a school where leaders were reported to be supportive, but they never had any direct communication with the project team, and there was no evidence of their involvement in the delivery/ organisation of IGR. T4 also described the programme as an opportunity for SATs and phonics pupil practice, hence misinterpreting the remedial nature of IGR.

In addition, the extent to which the IGR organisation model fitted well with existing arrangements in schools seemed to have an effect on the IGR implementation. Where IGR could be incorporated into the existing arrangements with few changes, then the programme implementation was smoother and there were few practical problems – this is, e.g., the case of T1 (and T5 from the next section). The implications of the IGR arrangements in such cases could be as limited as having to train non-IGR pupils to work independently. Yet, where more changes were needed then a few issues arose, especially in relation to the teacher/ TA time the other reading groups were getting during the year; this is the case of T4 and applies also to most teachers in the next section. There was also a third case represented by one school in this analysis that showed confidence in the IGR organisation and facilitated adjustments to enable teachers to use IGR with the suggested organisation, for instance by creating larger non-IGR groups. Yet, this was not associated with the same results for all teachers (the case of T2 and T7).

Mismatch between IGR fidelity and reading gains (T3, T5, T6 and T7)

These teachers' case studies present a mismatch between high fidelity and high reading gains and low fidelity and low reading gains which is inconsistent with the original hypothesis that using the IGR programme with good fidelity corresponds with higher reading gains for the IGR group. They all represent different combinations, T3: H/L; T5: L/MH; T6: M/L; and T7: M/L (M stands for *medium*). These teacher cases could be organised into two groups: the first has teachers with high/ medium fidelity scores and low reading gains (T3, T6 and T7); and the second involves the single case study of a teacher with low fidelity scores and relatively high gains (T5).

The first group of case studies involves teachers who despite their adherence to the programme procedures, taught groups that did not record good progress on the standardised assessments used. For T3, this was mainly due to her

mechanical approach to the programme implementation that became a barrier to using the programme strategies effectively and engaging pupils at a group level. T6 and T7 had both organisational difficulties related to their TAs and concerns about the effects of the delivery of IGR on the class pupils. T6 had a less experienced TA who was less comfortable to manage the rest of the class during teacher-led IGR, so the teacher had to pay attention to both her class and IGR group. T7, on the other hand, had a confident TA who took over the rest of the class during IGR, but this did not work well with her job-sharing teacher, resulting in tensions. As a result, T6 and T7 had a less smooth experience with the IGR organisation during the year. In addition to this, they were both concerned about the limited teacher time available for pupils in non-IGR groups (i.e. pupils not having the intervention). T6's IGR group had only 1 TA-led session a week (as opposed to 2), and this during a Friday assembly to give the TA time to read with other groups during the week – this intervention dosage problem could explain the low gains of this IGR group. T7, in turn, was reading with one of her non-IGR groups in assembly time, to make sure that all reading groups had reading time with the teacher every week. She also had to increase the number of pupils in the non-IGR reading groups (7-8 pupils per group), so that the IGR group could have only 4 pupils (this was due to this school having 34-5 pupils in each classroom) with implications for the teaching. These issues suggest that, although both teachers tried to implement the IGR model, this was only possible to a certain extent, as there was a less good fit between IGR and pre-existing arrangements (in relation to e.g. staffing and teaching arrangements). These issues might explain the low gains of their IGR groups despite the average or good fidelity of their IGR teaching.

On the other hand, T5 represents the case of a teacher with poor fidelity to the programme, but relatively high gains for her group. T5 received a low fidelity score mainly because aspects of the routine were missing as a result of her attempt to slow down the pace of the lesson using one book per week. Her decision was in response to the needs of her group; it is indicative that the aspects of the routine that were present were taught with good quality and attention to detail. The gains of her IGR group could be seen to suggest that good teaching decisions taken to address specific pupils' needs might be more important than adherence to the programme's protocol; or, in other words, that a programme needs to be flexible to adjust to particular circumstances. This teacher's case points also to issues relevant to the calculation of teaching fidelity in the study discussed below.

Fidelity was calculated from an index that was developed after the observations had taken place (and before knowing any statistical results). Since the programme team observation notes were made without strict standard procedures in mind (e.g. a scale with the main programme features), the observation data available were very varied in style and quality. In addition, there were some independent observations from the evaluation team but these were conducted with a focus on organisational decisions and challenges, so they could shed little light on the fidelity of teaching. Based on the programme team observations and using the fidelity index, the programme and evaluation team scores of the same teachers correlated highly (0.8), but the programme team scored all teachers consistently lower (a mean of 1.8 compared to 2.3). (The scores used for this analysis were the scores of the evaluation team.) This suggests that the scores should be seen as indicative of the quality of teaching rather than as accurate descriptors (index in appendix 7).

The fidelity index also had weaknesses, as it failed to distinguish between omissions and strategic changes in pace that were teacher responses to pupils' needs – with most notable example of alterations being the use of only one book per week. This resulted in lower fidelity scores for some teachers, as for example T5 (whose IGR group showed progress). A more sensitive instrument should be able to distinguish between good and less good teacher decisions, rather than expecting adherence to procedures that might not be relevant to the pupils at the time of the observation. Also, teachers who used the programme mechanically scored highly on the index, although they failed to engage in depth with the strategies involved, as in the case of T3. A better instrument should be able to distinguish between teachers who take a mechanical approach and teachers who engage with the programme and its strategies, especially the collaborative element of IGR.

An additional issue was that most teacher observations happened at three different intervals during the year, but for some teachers these happened earlier compared to others, when they were still experiencing teething problems. Since the programme team visits were meant to be supportive (i.e. they were not collecting evaluation data), they did not return for a final visit to 1-2 teachers about whose quality of teaching they were confident – although other forms of support were offered, e.g. emails and the supplying of additional materials. In some cases, this resulted in

lower fidelity scores that were reflective of their teaching difficulties at the beginning of the year (the case of T5 is relevant here as well).

14. Lessons learned about IGR (from programme team support visits)

A. IGR Training

1. **Literacy Adviser Induction/Training** – the setting up of expert local monitoring and support for teachers and TAs, making sure that Advisers are fully informed vis-à-vis IGR teaching practices and principles, and that we have developed a refined Lesson Observation Record, a common approach to its use, clear observation guidelines, and common advisory and TA training protocols
2. **Initial Teacher Training -** clarification of the task in hand, focusing on IGR as an addition to Wave One classroom early reading education for those children in need of intermediate
3. group reading teaching, and giving an in-depth and detailed overview of the principles and practices of IGR teaching-learning methodology including an introduction to teaching materials
4. **Early Local Follow-up Teacher Training** – early further clarification and training in a smaller local or school-based setting
5. so that teachers can re-visit and further embed their initial training in a practical, reflective way at an early stage
6. **Midway Local Follow-up Teacher Training** – addressing any concerns that may have arisen in practice over time and following through with further additional support and advice as appropriate.

B. Monitoring and Support Visits

1. It took time for teachers to adjust to the novel practice requirements of IGR teaching, and in particular to the fact that its aims and lesson protocols are distinct from and different to those of guided or other group reading/comprehension lessons. Guided reading lesson delivery routines are difficult to supersede – teachers are under a lot of pressure to follow government-recommended routines and practices, even where children are clearly unresponsive to them.

2. This was particularly true for new book introduction. Guided reading teaching protocol requires teachers to spend a large section of each lesson exploring features of books or texts with an emphasis on the use of questioning to develop children's analytic skills in relation to comprehension; in the context of Integrated Group Reading, which is about teaching struggling readers to read, and uses the (shorter) story familiarisation section of the lesson to inspire and engage them with story content and meaning, the guided reading metacognitive approach mitigates against rather than supports what is required. In combination with a misinterpretation of the purpose of the Lotto game, this can result in there being not enough time remaining in the lesson for good Collaborative Reading and Problem-Solving, something that is central in IGR to the development of the practice of reading itself for these same struggling readers.

3. **Year 2** teachers, especially those relatively young teachers who were struggling to work effectively with particularly hard-to-teach children, needed more targeted initial and follow-up IGR training programmes including the provision of and training in the use of a wider range of reading materials and accompanying games and activities at easier levels for their children. The need for more IGR materials at simple levels was evident from the Programme Team observations in both phases of the project. Initial training for Year 2 teachers would therefore include the following points:

4. Begin slowly with IGR – establishing a pace and routine that suits your children's learning - how to use one book over two lessons for best learning outcomes in the early stages (the Split Lesson) – finding and using supplementary IGR materials.

5. Make the most of your TA – secure overlearning at the level of the Word and using Win the Word to support the development of sequential listening and corresponding grapheme identification.

6. Practice orchestrating Collaborative Reading and Problem-Solving – the skill of working with group-based reading – harnessing children's individual strengths and weaknesses in the service of learning for all – keeping all of the children engaged in the reading task all of the time

7. Be aware of possible initial mismatch in many groups between book pitch and children's instructional learning requirements – a combination of an initial lack of assessment information for the programme developer/manager and teacher politeness or uncertainty meant that some IGR teachers and their children were left without proper guidance and correctly-levelled materials for too long.

8. There may be a lack of enough IGR materials at Turquoise level (and beyond) which by the beginning of the summer term were needed by many Year 3 groups (materials were produced for them using suitable books from other reading schemes, however). With Year 2 groups additional early materials are required.

9. Analysis of Lesson Observation Records revealed that experienced teachers in both Year 2 and Year 3 who understood IGR principles and made consistently good use of IGR lesson methodology at the Story Familiarisation and Lotto stages to pre-prepare their children for reading, also seemed to understand how to orchestrate Collaborative Reading and Problem-Solving successfully.

10. Lesson Observation analysis also indicated that the 26-week duration of the intervention was in many cases not long enough for younger, less experienced teachers to become expert across the full range of IGR lesson sections, most noticeably (not) in respect of IGR Collaborative Reading and Problem-Solving with a small group. This pointed to a need for more in-depth initial training of teachers in this regard, even though the same feature (a less strong delivery of Collaborative Reading and Problem-Solving) appeared also to be present for teachers in job shares who were delivering IGR on a once- rather than twice-a-week basis over the course of the project.

Summary

The Programme team looked for and supported where necessary the development of teacher expertise in relation to IGR understanding, and the exemplification of this in their delivery of classroom-based IGR lessons.

This involved looking at:

a. progress through IGR lesson sections,
b. depth of understanding in regard to each lesson section,
c. good lesson pace and appropriate timing of sections,
d. good preparation for Collaborative Reading and Problem-Solving in the Story familiarisation and Lotto sections,
e. a consistent IGR lesson routine over time,
f. the orchestration and understanding of good Collaborative Reading and Problem-Solving including good engagement and
g. detailed understanding of the children as learners,
h. thought and understanding in relation to the accurate matching of analytic phonics games to book levels and children's incremental progress in this regard,

Indications for possible IGR Programme Revisions

IGR teaching methodology and materials were developed over a number of years by the programme developer to match the learning requirements of readers who struggle. The methodology is based on several sound and well-tried pedagogic teaching and learning principles, and is designed intentionally, also, to provide a systematic structure within which children's need for narrative and play can be harnessed and celebrated in the service of their learning development. It is a sophisticated learning approach made up of several interconnected and interlocking parts that leads children from what is easier (the story) to what may be more difficult (words out of context and synthetic word building), but its use of games for both advance organising and overlearning, and the embedding of drawn response to story as the foundation of TA word-based follow-up work make it simple and nourishing for the children at the point of delivery.

In addition, over the course of programme team lesson observations of teachers using the methodology with their children over time, it has been clear that several possible refinements and additions would be welcomed. For instance,

a. The option of the insertion of a quick sight word revision check after Go Fish, just before the change to a new book, to make sure that children's new word knowledge has been retained overnight
b. More emphasis on the role and importance of the TA, especially with regard to ensuring that children have a secure grasp of Pelmanism Words (words out of context) before their next teacher-led lesson

c. The development of a full set of Win the Word activities (one per book pack) so that TAs (or teachers) can work consistently with children needing additional synthetic phonics training and practice using IGR materials
d. The development of clearer guidelines for both teachers and TAs
e. in relation to children's drawn and written IGR follow-up work
f. (this would be different for Years 2 and 3)
g. The development of clearer guidelines for Year 2 IGR teaching
h. More flexibility in some of the minor aspects of IGR lesson routine. For instance, teacher decision-making over when best to give their children a minute to look freely through their books (before Lotto or after Lotto), at what point to check children's grasp of what they have just read (before the SWAP game or at the very end of the lesson).
i. These things are always a matter of teacher judgment and intuition as the lesson is happening and lie outside the scope of recommended IGR lesson protocol

15. IGR interactions

Gender

There seem to be some statistically significant interactions between IGR and gender in Phase1, but not Phase 2. IGR had a more positive effect on Hodder scores for boys than for girls. For attitudes towards reading, IGR had a more negative effect on girls compared to boys.

	Dependent variable:									
	HodderDiff		SWRTDiff		hifamDiff		comDiff		attDiff	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment phase 1	-5.0**		-0.1		0.1		0.1		0.3**	
	(2.2)		(1.6)		(0.1)		(0.2)		(0.1)	
Treatment phase 2		1.5		-0.1		0.1		0.1		0.1
		(2.0)		(1.3)		(0.1)		(0.1)		(0.1)
Gender: girls	-2.9	-2.3	-1.1	-0.9	0.1	0.1	-0.3	-0.2	-0.02	-0.01
	(2.0)	(2.1)	(1.5)	(1.5)	(0.04)	(0.04)	(0.2)	(0.2)	(0.2)	(0.2)
Year: 3	1.1	-2.0	-3.0**	-4.1***	-0.003	-0.02	0.1	-0.1	0.05	0.01
	(1.8)	(1.7)	(1.4)	(1.0)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Treat Phase 1 * Girls	9.4***		-0.3		-0.1		0.1		-0.5**	
	(3.4)		(2.3)		(0.1)		(0.3)		(0.2)	
Treat Phase 2 * Girls		3.1		0.2		-0.03		0.1		-0.04
		(3.2)		(2.1)		(0.1)		(0.2)		(0.2)
Constant	3.8**	5.2***	5.3***	5.8***	-0.1**	-0.1**	0.1	0.2*	0.1	0.1
	(1.5)	(1.5)	(1.0)	(0.9)	(0.04)	(0.04)	(0.1)	(0.1)	(0.1)	(0.1)
Observations	198	215	212	226	212	226	209	225	209	225

Note: *p<0.1; **p<0.05;** p<0.01

Year Group

Most interaction effects are not statistically significant. In Phase 1, the IGR seemed to have a statistically significant positive effect on reading self-competence for Year 3 pupils, but not Year 2 pupils. This effect was not replicated in Phase 2.

	Dependent variable:									
	HodderDiff		SWRTDiff		hifamDiff		comDiff		attDiff	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment phase 1	-3.5		-2.0		-0.02		-0.1		0.004	
	(3.3)		(2.2)		(0.1)		(0.2)		(0.2)	
Treatment phase 2		4.4		-0.6		0.01		0.1		-0.1
		(2.7)		(1.6)		(0.1)		(0.2)		(0.2)
Year: 3	-0.6	-0.2	-4.8***	-4.9***	-0.1	-0.1	-0.2	-0.2	-0.1	-0.1
	(2.1)	(2.0)	(1.2)	(1.2)	(0.1)	(0.1)	(0.2)	(0.2)	(0.2)	(0.2)
Gender: girls	2.2	-0.7	-1.2	-0.7	0.01	0.1	-0.2	-0.2	-0.3**	-0.02
	(1.9)	(1.7)	(1.1)	(1.1)	(0.04)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Treat Phase 1 * Year 3	3.4		3.8		0.1		0.6**		0.3	
	(3.7)		(2.7)		(0.1)		(0.2)		(0.2)	
Treat Phase 2 * Year 3		-3.8		1.3		0.1		0.1		0.3
		(3.4)		(1.9)		(0.1)		(0.2)		(0.2)
Constant	3.3*	3.9**	6.2***	6.1***	-0.04	-0.05	0.2*	0.2*	0.2*	0.2
	(1.7)	(1.7)	(1.0)	(1.0)	(0.03)	(0.03)	(0.1)	(0.1)	(0.1)	(0.1)
Observations	198	215	212	226	212	226	209	225	209	225

Note: *p<0.1; **p<0.05;** p<0.01

English as additional language (EAL)

The IGR effect on Hodder scores was statistically significant for EAL pupils in Phase 1, but not in Phase 2.

	Dependent variable:									
	HodderDiff		SWRTDiff		hifamDiff		comDiff		attDiff	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment phase 1	-2.8		-0.2		0.03		0.1		0.1	
	(1.8)		(1.4)		(0.1)		(0.1)		(0.1)	
Treatment phase 2		1.9		0.03		0.02		0.1		0.1

		(1.9)		(1.1)		(0.1)		(0.1)		(0.1)
EAL	0.3	0.3	2.9	2.9	-0.1	-0.1	0.2	0.2	0.3	0.3
	(2.9)	(2.9)	(2.7)	(2.7)	(0.1)	(0.1)	(0.2)	(0.2)	(0.2)	(0.2)
Treat Phase 1 * EAL	7.4*		-1.5		0.1		-0.2		-0.2	
	(4.4)		(4.4)		(0.1)		(0.3)		(0.3)	
Treat Phase 2 * EAL		2.4		-1.5		0.2*		-0.2		-0.2
		(3.9)		(3.7)		(0.1)		(0.3)		(0.3)
Constant	3.5***	3.5***	3.1***	3.1***	-0.1	-0.1	0.1	0.1	0.1	0.1
	(1.2)	(1.2)	(0.7)	(0.7)	(0.03)	(0.03)	(0.1)	(0.1)	(0.1)	(0.1)
Observations	198	215	212	226	212	226	209	225	209	225

Note: *p<0.1;* ***p<0.05;*** p<0.01

Special Educational Needs

In Phase 2 (but not in Phase 1) IGR had a more positive effect on reading self-competence and attitude for children with SEN.

	Dependent variable:									
	HodderDiff		SWRTDiff		hifamDiff		comDiff		attDiff	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment phase 1	-0.9		-1.6		0.1		0.01		-0.1	
	(3.1)		(1.8)		(0.1)		(0.2)		(0.2)	
Treatment phase 2		0.5		0.3		0.1		-0.2		-0.2
		(2.7)		(1.5)		(0.1)		(0.2)		(0.2)
SEN	-2.2	-2.2	-0.4	-0.4	-0.001	-0.001	-0.3*	-0.3*	-0.3	-0.3
	(2.1)	(2.1)	(1.2)	(1.2)	(0.05)	(0.05)	(0.2)	(0.2)	(0.2)	(0.2)
Treat Phase 1 * SEN	-2.7		2.9		-0.1		0.1		0.3	
	(3.9)		(2.4)		(0.1)		(0.3)		(0.3)	
Treat Phase 2 * SEN		3.5		-1.0		0.02		0.5**		0.5**
		(3.1)		(1.9)		(0.1)		(0.2)		(0.2)
Constant	5.0**	4.9**	3.9***	3.9***	-0.1*	-0.1*	0.3**	0.3**	0.3**	0.3**
	(2.0)	(2.0)	(1.0)	(1.0)	(0.03)	(0.03)	(0.1)	(0.1)	(0.1)	(0.1)
Observations	198	215	212	226	212	226	209	225	209	225

Note: *p<0.1;* ***p<0.05;*** p<0.01

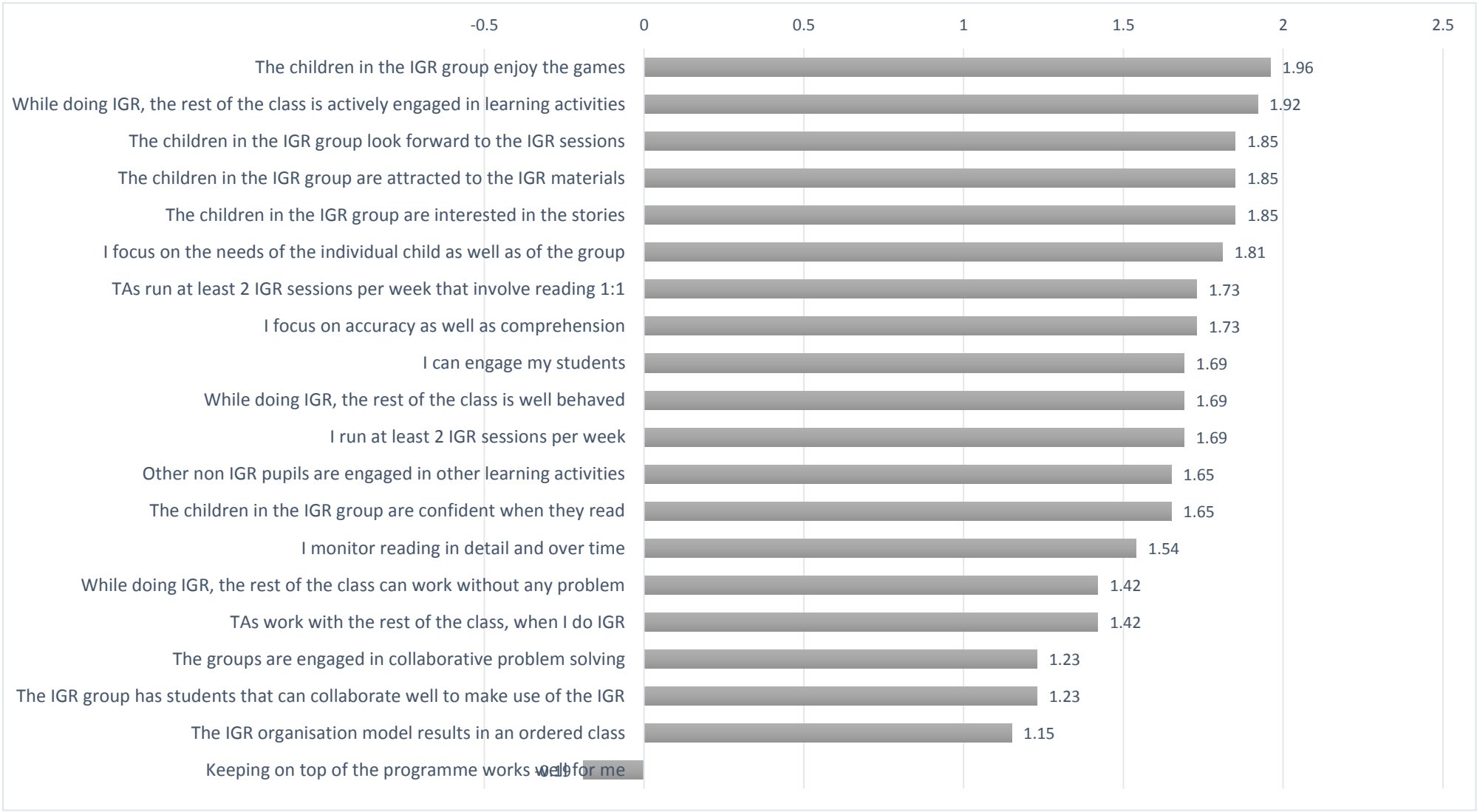
Pupil Premium

In Phase 2 (but not in Phase 1) IGR had a more positive effect on SWRT for children with Pupil Premium.

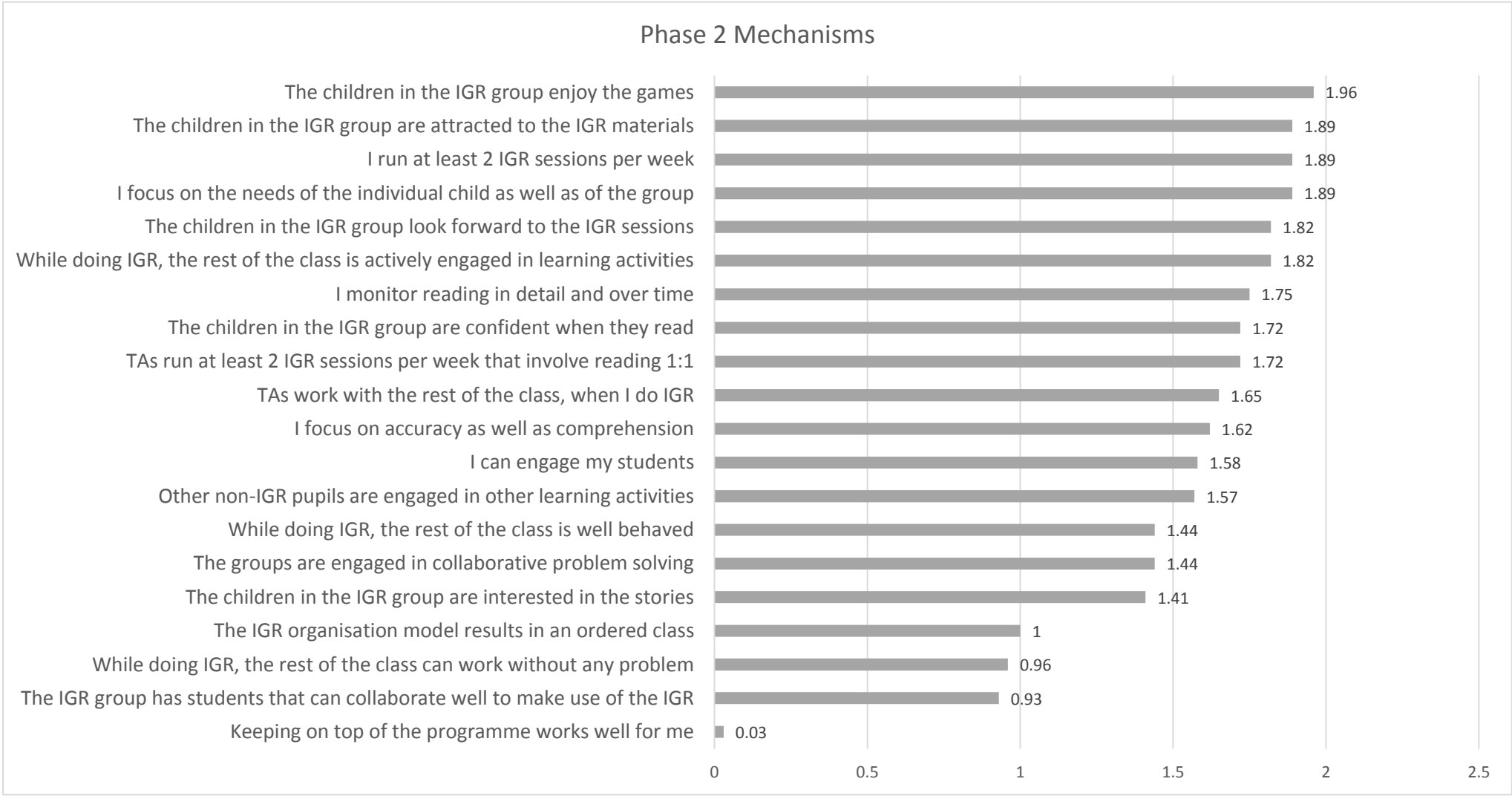
	Dependent variable:									
	HodderDiff		SWRTDiff		hifamDiff		comDiff		attDiff	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment phase 1	-0.8		-1.4		0.1		0.1		0.2*	
	(1.8)		(1.5)		(0.1)		(0.1)		(0.1)	
Treatment phase 2		2.2		-1.5		0.03		0.1		0.1
		(1.9)		(1.2)		(0.1)		(0.1)		(0.1)
Pupil Premium	-1.0	-1.0	-2.2	-2.2	-0.1	-0.1	0.04	0.04	-0.02	-0.02
	(2.0)	(2.0)	(1.8)	(1.8)	(0.1)	(0.1)	(0.2)	(0.2)	(0.2)	(0.2)
Treat Phase 1 * Pupil Premium	-2.4		3.7		-0.1		0.01		-0.4	
	(3.2)		(3.0)		(0.1)		(0.3)		(0.3)	
Treat Phase 2 * Pupil Premium		1.1		4.6**		0.1		-0.01		-0.1
		(3.3)		(2.2)		(0.1)		(0.3)		(0.2)
Constant	3.9***	3.9***	4.2***	4.2***	-0.1	-0.1	0.1	0.1	0.1*	0.1*
	(1.1)	(1.1)	(0.9)	(0.9)	(0.04)	(0.04)	(0.1)	(0.1)	(0.1)	(0.1)
Observations	198	215	212	226	212	226	209	225	209	225

Note: *p<0.1;* ***p<0.05;*** p<0.01

16. Phase 1 mechanisms (from CMO questionnaire)



17. Phase 2 mechanisms (from CMO questionnaire)



18. CMO (Contexts, Mechanisms and Outcomes) questionnaire

1. What are the end of phase 1 outcomes of the IGR programme?

Please identify a position on the scale (2, 1, 0) either closer to the positive or negative expression of the same item. Zero indicates that neither position is relevant.

Outcomes	Closer to this side		Neither		Closer to this side	
IGR children have made overall reading progress	2	1	0	1	2	There is no indication that IGR children have made overall reading progress
IGR children have made progress in their oral language	2	1	0	1	2	There is no indication that IGR children have made progress in their oral language
IGR children have developed a more positive attitude to reading	2	1	0	1	2	There is no indication that IGR children have developed a more positive attitude to reading
IGR children have developed a more positive attitude to school	2	1	0	1	2	There is no indication that IGR children have developed a more positive attitude to school
Some IGR pupils are concerned being visible in a low attainment reading group	2	1	0	1	2	IGR pupils aren't concerned being visible in a low attainment reading group
Other non IGR children have made overall reading progress	2	1	0	1	2	There is no indication that other non IGR children have made overall reading progress
I have developed my literacy teaching content knowledge	2	1	0	1	2	There is no indication that I have developed my literacy teaching content knowledge
I am confident to use the IGR organisation as an inclusive model of support	2	1	0	1	2	I am less confident to use the IGR organisation as an inclusive model of support
I can organise better my cooperation with TAs	2	1	0	1	2	I still have difficulties in cooperating with TAs
I would like to be less responsible for inclusive teaching	2	1	0	1	2	I am happy to be responsible for inclusive teaching

Would you like to suggest any other outcomes that you see as relevant?

2. Select now the reasons that you think are relevant to produce these outcomes:

Please identify a position on the scale (2, 1, 0) either closer to the positive or negative expression of the same item. Zero indicates that neither position is relevant.

Reasons	Closer to this side		Neither		Closer to this side
I have difficulties in engaging my students	2	1	0	1	2	I can engage my students
I monitor reading in detail and over time	2	1	0	1	2	I do not always monitor reading in detail and over time
The groups are engaged in collaborative problem solving	2	1	0	1	2	The groups are not always engaged in collaborative problem solving
I focus on accuracy as well as comprehension	2	1	0	1	2	I focus more on one aspect of reading
I focus on the needs of the individual child as well as of the group	2	1	0	1	2	I focus more either on group or individual needs
I run at least 2 IGR sessions per week	2	1	0	1	2	I run fewer than 2 IGR sessions per week
TAs run at least 2 IGR sessions per week that involve reading 1:1	2	1	0	1	2	TAs run less than 2 IGR sessions per week that involve reading 1:1
TAs work with the rest of the class, when I do IGR	2	1	0	1	2	There is no TA support for the rest of the class, when I do IGR
While doing IGR, it is not clear what the rest of the class has to do	2	1	0	1	2	While doing IGR, the rest of the class is actively engaged in learning activities
While doing IGR, the rest of the class is unsettled	2	1	0	1	2	While doing IGR, the rest of the class is well behaved
The IGR group has students that have difficulties in collaborating well to make use of the IGR	2	1	0	1	2	The IGR group has students that can collaborate well to make use of the IGR
The IGR organisation model results in an unsettled class	2	1	0	1	2	The IGR organisation model results in an ordered class
While doing IGR, the rest of the class can work without any problem	2	1	0	1	2	While doing IGR, the rest of the class is a problem for the teachers
Keeping on top of the programme is demanding for me	2	1	0	1	2	Keeping on top of the programme works well for me
The children in the IGR group are interested in the stories	2	1	0	1	2	The children in the IGR group are not always interested in the stories
The children in the IGR group enjoy the games	2	1	0	1	2	The children in the IGR group are not always interested in the games
The children in the IGR group are attracted to the IGR materials	2	1	0	1	2	The children in the IGR group are not always interested in the IGR materials
The children in the IGR group look forward to the IGR sessions	2	1	0	1	2	The children in the IGR group are not always interested to the IGR sessions
The children in the IGR group are confident when they read	2	1	0	1	2	The children in the IGR group are less confident when they reading
Other non IGR pupils are engaged in other learning activities	2	1	0	1	2	Other non IGR pupils are not always engaged in learning and are unsettled

3. What are the contexts that enable the reasons that you see as relevant to produce the end of phase 1 outcomes of IGR?

Please identify a position on the scale (2, 1, 0) either closer to the positive or negative expression of the same item. Zero indicates that neither position is relevant.

Context factors	Closer to this side		Neither		Closer to this side	
School leaders are supportive	2	1	0	1	2	School leaders are unsupportive
The school has few available TAs	2	1	0	1	2	The school has TAs to support me
The school struggles to release me for training or other activities	2	1	0	1	2	The school is happy to release me for training or other activities
The literacy advisor has the necessary time to support me and the participating TAs	2	1	0	1	2	The literacy advisor struggles to find time to support me and the participating TAs
The programme team provides relevant and timely support	2	1	0	1	2	The programme team does not provide relevant and timely support
The programme training covers every area of the programme implementation	2	1	0	1	2	The programme training does not cover every area of the programme implementation
I am less confident in literacy teaching	2	1	0	1	2	I am confident in literacy teaching
I am interested in supporting struggling readers myself	2	1	0	1	2	I am unwilling to support struggling readers myself
There are co-operation difficulties between the TAs and me	2	1	0	1	2	TAs have good co-operation with me
There is strong overall school commitment to literacy provision	2	1	0	1	2	The overall school commitment to literacy provision is weak and vague

Would you like to suggest any other reasons or context factors that you see as relevant?

Name:

Date:

19. Fidelity index for IGR teaching

For each IGR teacher, the support visit notes (JS and local advisors) to be read and the following ratings to be made on the 9 dimensions. There is a 4 point scale and a single rating is to be given for each dimension. The poles of the scale are anchored by more specific descriptors of what a 3 and a 0 mean. Judgement is required in making each rating taking account of all descriptors after balancing indicators that might be contrary.				
1. Relaxing into learning – pace, organisation, atmosphere	3	2	1	0
	Classroom organisation is the suggested one			Classroom is unsettled or organisation is different from the suggested
	Teacher works closely with children with high quality, detailed teaching			Teacher not fully engaged with the children
	Teacher moves seamlessly from one section to the next			
	Teacher and pupils relaxed and enjoying lesson			Unsettled and bored
	Close-knit group and focused			Disorganised and unfocused
	Teacher understands IGR teaching rationale			
	Appropriate pacing and timing for lesson sections; balance between sections			
2. Sequence	3	2	1	0
	Full sequence	Miss 1 step	Miss 2 steps	Misses 3 steps in the sequence
3. IGR lesson start	3	2	1	0
	At lesson start: drawing & sentences acknowledged			At lesson start: drawing & sentences not acknowledged
	Recall/comprehension check from previous book			Recall/comprehension check missing or perfunctory
	Go fish is played correctly			Go fish not played correctly (no bonus turns for remembering, goes on too long).
4. Everyone well prepared and engaged	3	2	1	0
	Teacher knows stories well			
	Teacher ensures children are engaged in story before they read			
	Children prepared – familiar with story and Lotto before reading			
	Teacher knows how to story tell, using illustrations to engage children; enabling children to understand higher-level stories allowing spontaneous contributions			Employs guided reading protocols, e.g. excessive questioning (asking for inferences/ predictions), requesting meta-cognitive comprehension analysis or interrupting to make teaching points. Teacher misunderstands the meaning of storytelling – reads the text itself to the children

5. Using Lotto properly	3	2	1	0
	Make explicit links to stories			Not linking to stories
	Elucidate word meanings for phonological to visual mapping			Failing to elucidate word meanings
				Asking children to read words on lotto board at start
				Showing word cards to children Using lotto to discuss grammar
				Keeping children at basic level phonological-visual mapping
6. Collaborative reading and problem solving	3	2	1	0
	Leaves enough time for this section			Not enough time
	Individual children enough time for problem-solving (e.g. use of phonic skills)			Children not do very much reading No time for individual problem solving; teacher racing through it themselves; reading the whole text chorally
	Teacher understands strategic use of choral reading (e.g. to support re-reading for fluency)			
	Teacher checks children's comprehension with 1-2 good questions			Teacher forgets to check recall and comprehension at end of reading, or while dealing SWAP cards.
7. SWAP game at appropriate level	3	2	1	0
	Teacher selects a hard enough, well-pitched game for the children.			Teacher chooses games that are too easy or too hard; not leaving enough time for SWAP and / or using SWAP time for other phonics work.
8. Ensuring good TA work	3	2	1	0
	Teacher remembers to prepare children for their TA session			Teacher not leaving enough time to prepare children for the TA follow up work
	Detailed record-keeping between teacher and TA via Daily Record			Teacher not initiating work children will do with TA