



UNIVERSITY OF
Southampton



Foreign Language Education: Unlocking Reading (FLEUR)

A study into the teaching of reading to beginner learners of French in secondary school

Dr Robert Woore, University of Oxford Department of Education; Professor Suzanne Graham, University of Reading, Institute of Education; Dr Louise Courtney, University of Reading, Institute of Education; Dr Alison Porter, University of Southampton, Department of Modern Languages and Linguistics; Clare Savory, University of Oxford Department of Education

Contents

Acknowledgments	4
Foreign Language Education: Unlocking Reading (FLEUR)	5
Executive summary.....	5
Project aims	5
Research questions.....	5
Methods.....	5
Key findings.....	6
Key recommendations.....	7
1. Background	9
Motivation, attainment and uptake in secondary MFL.....	9
Reading in the secondary MFL curriculum.	9
Strategic reading.....	10
Phonics in French.....	11
2. Aims of our study.....	12
3. Methods.....	13
Overview.....	13
Outcome measures.....	13
Allocation of classes to experimental groups.....	14
Sample: participating schools.....	15
Interventions: the programmes of instruction	16
Responsive teaching and ‘fidelity to condition’	19
Attrition	20
Analysis	20
Ethical issues.....	21
3. Findings.....	21
Reading comprehension	21
Phonological decoding.....	23
Vocabulary knowledge	24
Self-efficacy for French reading.....	26
Motivation	29
Strategic behaviour.....	31
Students’ views on the programmes of instruction	33
Teachers’ views on the programmes of instruction.....	36
Discussion and conclusions	37
Impact on reading comprehension	37
Other linguistic outcomes: phonological decoding and vocabulary knowledge.....	39
Strategic behaviour.....	40

Attitudinal outcomes: self-efficacy and motivation	40
Other views on the interventions.....	40
Key recommendations.....	41
References	43
Appendix.....	45
Appendix 1: Consort flow diagram of study design and sample	45
Appendix 2: Details of statistical tests.....	46

Acknowledgments

We are very grateful to the Nuffield Foundation for funding the FLEUR project, and in particular to Cheryl Lloyd, Hannah Broad, Kim Woodruff and Josh Hillman for being so approachable and supportive.

The FLEUR study would not have been possible without the support and participation of 36 Modern Foreign Languages teachers in secondary schools around the country. In order to preserve anonymity, we cannot name the teachers or schools who took part. Nonetheless, we hope they know how grateful we are for all of the time, thought and effort that they put into the project. We are also grateful to the head teachers and heads of department who gave their permission for the study to be conducted in their schools; and, of course, to the many students who voluntarily completed our various tests and questionnaires.

We were assisted in visiting schools, administering tests, marking tests and inputting data by a large team of additional research assistants, whose help and professionalism are hugely appreciated: Marwa Ahmed, James Turner, James Corbet, Coralie Clerc, Rachel Parker, Monika Maendler-Zhao, Anna Niedbala, Amy Wallington, Amber Dudley, Olivia Clarke, Esther Outram, Zening Yang, Tom Graham and Claudia Pope, Beatriz Melendro Martinez, Bethan Thomson, Heike Kruesemann and Julia Hofweber.

We would also like to say thank you to all of the following:

- Our Modern Languages Expert Advisory Group: Kirsten Gregory, Anna Cross, Geoff Roberts, Patricia Driscoll, Cat Draper and Alison Seath.
- Our methodological advisors, both formal and informal: Professor Kathy Sylva, Professor Steve Strand, Dr Josh McGrane, Professor Jim Milton, Fiona Jelley and Sandra Mathers.
- The Association for Language Learning and Hannah Brown for helping us publicize our end-of-project conference.
- David Griffiths for his invaluable advice on managing the project
- Jenny Lawlor, Louisa Summers and Claire Stevens for managing the project finances and contracts

Finally, we owe a large debt of gratitude for the tireless efforts of our main Research Assistant and Project Manager, Clare Savory. Clare went above and beyond the call of duty in her devotion to the FLEUR project and she was absolutely crucial to its smooth running and successful completion.



The Nuffield Foundation is an endowed charitable trust that aims to improve social well-being in the widest sense. It funds research and innovation in education and social policy and also works to build capacity in education, science and social science research. The Nuffield Foundation has funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. More information is available at www.nuffieldfoundation.org

Foreign Language Education: Unlocking Reading (FLEUR)

A study into the teaching of reading to beginner learners of French in Year 7, funded by the Nuffield Foundation
Robert Woore, Suzanne Graham, Alison Porter, Louise Courtney and Clare Savory

Executive summary

Project aims

A sense of making progress at the start of Year 7 is crucial to sustaining motivation and hence enabling further progress in Modern Foreign Languages (Graham, Courtney, Tonkyn & Marinis, 2016). In particular, learners need to rapidly develop the foundation literacy skills essential at Key Stage 3 and 4, as well as a sense of self-efficacy or confidence to enable them to persevere with language learning. Our study therefore aimed to strengthen the evidence base on which teachers can draw when teaching reading in Year 7 Modern Foreign Languages (MFL), with a focus on French as the most widely taught language in England at the start of secondary school.

We explored two approaches to teaching reading in MFL, namely explicit phonics instruction and explicit reading strategy instruction. Phonics instruction means teaching students about the relationships between the written symbols of the language and the spoken sounds they represent, helping them to ‘sound out’ written words in order to be able to pronounce them accurately. Reading strategy instruction means teaching students how to use particular strategies to help them comprehend written text more effectively – for example, how to deduce the meaning of an unfamiliar word based on its context in a written passage. We also investigated whether any benefits of such explicit instruction was actually due to that instruction itself, or whether the same outcomes could be achieved simply through giving students practice in tackling more challenging reading texts. We aimed to investigate what impact these three approaches (Strategies, Phonics and Texts only) might have on learners’ reading comprehension, their ability to match the sounds of words to their written form (phonological decoding), their vocabulary knowledge, their strategic behaviour while reading French, their self-efficacy or confidence in reading French, and their broader motivation for learning the language. Finally, we also wished to know how learners and teachers felt about the kinds of instruction used.

Research questions

1. To what extent does a programme of explicit instruction in (a) reading strategies or (b) phonics lead to improvements in French reading proficiency amongst Year 7 MFL students, as compared to a programme of reading challenging texts?
2. To what extent do any of the three programmes of instruction lead to improvements in: (a) students’ French phonological decoding proficiency; (b) their strategic behaviour when reading in French; (c) their vocabulary knowledge; and (d) their motivation for learning the language?
3. Is one of the programmes of instruction more effective than the others on any of the above measures?
4. What are teachers’ and students’ views on the three programmes of instruction?

Methods

Participants

Thirty-six intact Year 7 classes (approximately 900 students in total), drawn from a diverse range of state schools, took part in the study. Year 7 is the first year of secondary school, with students aged 11-12 years. The classes were allocated to three groups: those using phonics instruction, strategy instruction or texts only. To allocate the classes, a process called ‘minimization’ was used. This is an alternative to random allocation recommended for studies like ours with samples of 100 schools or fewer (for more details see below).

Instruction

Over 16 weeks, from January 2017 to Summer 2017, teachers were asked to spend 20-30 minutes per week on the intervention teaching, and to refrain from giving explicit instruction pertaining to the other experimental groups during this period.

All groups worked with a series of eight 'pedagogical texts' specially created for the project. These covered topics related to francophone culture and were intended to be linguistically challenging (but not to the point of being inaccessible) for Key Stage 3 MFL students. Learners in the Strategies group were taught 8 strategies to help them understand the meaning of the texts. Those in the Phonics group were taught a series of key French grapheme-phoneme correspondences, and then used the pedagogical texts to practise these, reading the texts aloud. They also answered simple, factual questions along with some more open-ended, reflective questions about what they had read, but were not taught any strategies to understand the texts. Learners in the Texts group received no explicit instruction in either strategies or phonics. They answered the same questions on the pedagogical texts as the Phonics group.

Outcome measures

Learners were assessed at three time points – immediately before, immediately after and six months after experiencing the instruction (henceforth times 1, 2 and 3). Despite our best efforts to maintain the sample size throughout the project, we experienced considerable attrition at time 3. We lost three schools as well as a number of individual students in other schools. The most frequent reasons for this were where the participating teacher moved schools and their replacement did not wish to participate in the project; participating classes were re-organized due to 'setting' by attainment or other changes; or they had switched from studying French to Spanish, and so were no longer considered eligible to participate. Overall, our sample decreased from around 900 students at times 1-2 to 600 at time 3. For this reason, we have lower confidence in our findings at time 3. Our principal analyses are therefore based on the data from times 1 and 2.

Reading comprehension was assessed through translation tasks and comprehension questions; phonological decoding through a specially-developed 'Sound-Alike Task' (SALT), in which learners tackled a series of items requiring them to identify which of three written words sounded the same when read aloud. Vocabulary was assessed through a simple yes/no task, where they indicated whether they knew certain French words or not.

After completing the reading and SALT tasks, learners were asked to indicate their level of self-efficacy for completing similar tasks in the future. Information about motivation, strategic behaviour and views on the programmes of instruction was captured from students and teachers through short questionnaires. We also conducted interviews with 18 students and 14 teachers, to gather their views in greater depth.

Data analysis

We analysed our data using a range of statistical techniques, to examine whether there were any 'statistically significant' differences between the groups' scores and responses at each time point – in other words, results which are highly unlikely to have occurred unless they reflected a genuine difference between the groups, such as if one of the programmes of instruction was more effective in promoting reading comprehension than the others. For many of our analyses, we used multilevel models to take account of teacher and school effects (i.e. the fact that being in the same class, with the same teacher, in the same school is likely to influence students' outcomes). We were also able to control for students' prior academic attainment in many analyses, although this data was unfortunately not available for our whole sample.

Key findings

- On average, students in all three groups made statistically significant progress in **reading comprehension** over the course of the intervention. There was then a levelling off of progress between times 2 and 3. We found no evidence that any of the three programmes of instruction was more effective than the others in improving students' reading outcomes, either in the immediate or the longer term.
- Students in all three groups made significant progress in **phonological decoding**, though for the Strategies and Texts group, the amount of progress was small. There was a significant advantage for the Phonics group in terms of progress in decoding over the course of the intervention. Between times 2 and 3, there were no significant changes in any group's decoding scores.
- Students in all three groups also showed significant increases in **vocabulary knowledge** between times 1 and 2. Between times 2 and 3, vocabulary knowledge changed only very slightly or not at all. There was evidence

of an advantage for the Phonics group over the Texts groups in the vocabulary gains made over the course of the intervention, and for the Strategies group also when prior attainment was controlled for. Hence the advantage was clearest for the Phonics group, which showed the largest increases. Taken together with other recent evidence, this suggests that phonics instruction may be beneficial for vocabulary learning in an alphabetic foreign language.

- There were few differences in the **strategic behaviour** of the three groups before the intervention, although the Texts group appeared to use more strategies at time 1, presenting a possible confounding variable. At time 2, there were several areas of significant difference in strategic behaviour between (a) the Strategies and the Phonics groups and (b) the Texts and Phonics groups, with evidence of most strategy development in the Strategies group. There was some evidence of regression in strategic behaviour at time 3.
- In terms of **overall self-efficacy for reading**, all learners became more confident over the course of the intervention. Self-efficacy then declined between times 2 and 3, but remained above time 1 levels. On a descriptive level, the Strategies group showed the most improvement between times 1 and 2, and displayed the most widespread gains for individual aspects of self-efficacy. However, no statistically significant differences between the groups were found in terms of overall self-efficacy at any of the time points for the main analysis, though there was a significant advantage for the Strategies group over the Texts group between times 1 and 2 when prior attainment was controlled for.
- In the sample as a whole, students started Year 7 with a high level of **positivity for French**, which then declined at time 2 and again at Time 3, echoing the findings of previous research. No significant differences were found between the groups in terms of positivity towards French. The same was also true for learners' **sense of how good they were at French**.
- In terms of **participants' views on the interventions**, both learners and teachers in all groups showed a high level of positivity towards the programmes of instruction and the pedagogical texts. In particular, they seemed to enjoy the cultural content of the texts, and considered them to offer rich opportunities for developing French reading skills.

Key recommendations

Based on the project's findings and bringing these together with previous research evidence in the area, we offer the following key recommendations for MFL teachers in relation to the teaching of French reading to beginner learners:

- **Expectations regarding the kinds of written texts that beginner learners of French can tackle should be raised.** Our study shows that it is possible to use challenging, engaging texts, covering cultural topics of genuine interest to students, even with beginner learners. Further, our study suggests that the use of such texts can promote several aspects of reading and language development, as well as being popular with both teachers and learners. Therefore, we believe that including such texts in the Year 7 curriculum will be beneficial for students' linguistic and motivational development. They should not be restricted to a sole diet of the shorter, simpler and more predictable kinds of texts that are traditionally associated with this age group.
- **An integrated approach to French reading instruction – combining explicit instruction in both Strategies and Phonics with the use of appropriately challenging, engaging texts – is more likely to be beneficial than any of these approaches in isolation.** All three groups in our study made significant progress in reading French. However, there was evidence of an advantage for the Phonics group in phonological decoding, and for both the Strategies and the Phonics groups (most clearly for the latter) in vocabulary learning. There was also some evidence that the Strategies group made more progress in strategic behaviour and self-efficacy. Bringing these benefits together, alongside the use of challenging texts, is likely to lead to more engaged learners who have the linguistic knowledge, strategic proficiency, and resilience to comprehend written texts and hence benefit from them.
- **These approaches need to be used in the long term**, beyond Year 7, to retain their benefits, and they should be integrated into a focus on overall literacy development.
- We consider phonological decoding (the ability to read words aloud accurately and fluently) to be a foundation literacy skill in a foreign language. It cannot be assumed that students will develop this skill

spontaneously in French. **It is highly likely that explicit phonics instruction is beneficial, and indeed may be necessary, for many MFL students to learn to decode in French.**

- **Further development work should be undertaken with teachers to enable them to adapt their existing Schemes of Work,** to incorporate an integrated approach to reading instruction for learners with a range of different prior attainment levels, across a range of year groups and in a range of languages.

1. Background

Motivation, attainment and uptake in secondary MFL

The government has recently reaffirmed the importance of Modern Foreign Languages (MFL) in the school curriculum, reflecting the strategic importance of the subject for both individuals and wider society. Being able to speak a foreign language has instrumental value in terms of travel, employment (CBI, 2012) and university entry (e.g. Sammons, Toth & Sylva, 2015). It can also bring wider cognitive benefits (Bialystok, 2007) and increased intercultural understanding: in the words of the new National Curriculum (DfE, 2013), language learning offers “a liberation from insularity and provides an opening to other cultures”.

Nonetheless, MFL faces persistent problems of poor student attainment and motivation. A recent study ranked England bottom in Europe in terms of students’ foreign language attainment by age 14 (European Commission, 2012). Low motivation for language learning amongst secondary school students is also widely documented, with a particular ‘dip’ towards the end of Year 7 and in Year 8 (e.g. Coleman et al., 2007).

This dip in motivation in early secondary school has been linked to students’ lack of a sense that they are making any real progress in the subject (Macaro, 2008). In a recent investigation into primary/secondary transition in MFL, also funded by the Nuffield Foundation, Graham et al. (2016) likewise found that “a sense of progress” (p. 682) was a key motivator for students in Years 6-7. In Graham et al.’s study, students’ motivation actually increased as they moved from primary school into Year 7, with one main factor reported to be the greater progress they felt they were now making in French. By the end of Year 7, however, their motivation already appeared to be waning, as they seemed to be losing “the optimism that competences can be developed” (Spinath & Steinmayr, 2008, p. 1567, in Graham et al., 2016, p. 698).

In turn, these problems of motivation have been reflected in sharp and sustained declines in the numbers of students opting to study languages at Key Stage 4 (Board & Tinsley, 2015), with knock-on effects of poor recruitment at sixth form and degree level (Tinsley & Doležal, 2018). Thus, an urgent challenge for researchers and teachers is to develop effective pedagogical approaches to promote and sustain students’ progress and motivation as they move into, and progress through, Key Stage 3. This might in turn give them more solid foundations and greater confidence for studying the subject at Key Stage 4. The challenge is particularly acute given the limited curriculum time available for MFL learning in secondary school (usually around 2 to 3 hours per week).

Reading in the secondary MFL curriculum.

Foreign language reading can make a crucial contribution to wider MFL learning, for reasons including the following:

- As one of the four main language skills, it forms part of students’ broader communicative competence.
- Reading provides linguistic ‘input’, allowing students to encounter new language and consolidate what they know (e.g. in terms of vocabulary and grammatical structures).
- It supports autonomous learning, particularly outside the classroom. Students can read foreign language material on the internet, in textbooks and as an integral part of homework tasks.
- It offers a window on the target language culture – not only through books but also websites and blogs, song lyrics, youtube video captions, social messaging, and so forth.
- Foreign language reading may impact positively on English literacy, for example through the development of transferable reading strategies and knowledge of phonic decoding (Murphy et al., 2014).

Rightly in our view, the new National Curriculum (DfE, 2013) sets ambitious goals for MFL reading instruction at both KS2 and KS3, including the use of literary texts such as stories, songs and poems. Likewise, our own research-informed principle to underpin the teaching of reading – formulated as part of our Professional Development Consortium in MFL (PDCinMFL: see Macaro, Graham & Woore, 2015) – advocates the use of “a greater range of

more challenging texts” containing considerable proportions of unfamiliar language beyond students’ current productive level.

However, current approaches to reading instruction often appear to be rather limited. Various studies, as well as anecdotal evidence and our own classroom observations, suggest that many Key Stage 3 students experience a very restricted diet of reading materials. In contrast to the ambitions of the new National Curriculum, texts are often composed of mainly simple, familiar language on trivial or unadventurous topics (Andon & Wingate, 2012; McGowan and Turner, 1994; Ofsted, 2011). We would argue that it is little wonder students lose interest in reading in the foreign language. The subject matter does not match their cognitive level.

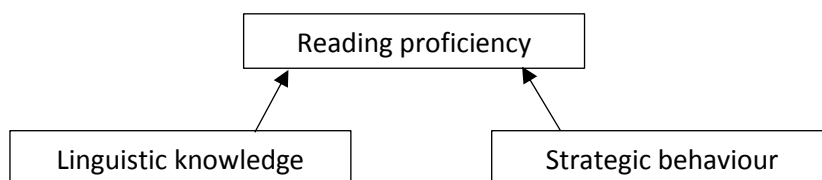
Strategic reading

If students are to access more challenging texts, it is likely that they will need to deploy appropriate strategic behaviour to compensate for gaps in their current linguistic knowledge. Examples of strategic reading behaviour might be:

- Using context to infer the meanings of unknown words
- Thinking about whether your initial understanding of a word or sentence makes sense in the wider context, or in light of new information as you read on
- Reminding yourself not to give up
- Sounding out an unfamiliar word to see this helps you recognize it

In other words, we can see reading outcomes (i.e. what students understand from a given written text) as a product of linguistic knowledge on the one hand – the vocabulary and grammatical structures that learners know – and their strategic behaviour on the other (Figure 1). The balance between the two will depend on the text itself and how much unfamiliar language it contains. A harder text, containing more unfamiliar words and structures, will require more in the way of compensatory strategic behaviour.

Figure 1: the contributions of linguistic knowledge and strategic behaviour to reading proficiency



There is evidence from research conducted in English secondary schools that Year 7 learners of French generally do not develop such effective strategic behaviour for foreign language reading spontaneously (Erler, 2003). Instead, Erler found, many students when completing reading exercises in class tended to report a reliance on strategies such as ‘wait for the teacher to explain’, ‘ask a friend what it means’, ‘ask the teacher what it means’ and ‘invent a meaning’ – in other words, strategies that involve disengaging from the text itself.

This raises a key question for teachers, and one that is hotly contested in both the research and pedagogical literature: can explicit instruction in reading strategies help learners become effective readers? Some argue that the limited available lesson time would be better spent teaching the language itself, rather than teaching strategies to compensate for insufficient language knowledge (e.g. Swan, 2008). On the other hand, as Woore (2014) argues, in the two or three hours of teaching per week typically allocated to MFL in Key Stage 3, it would take a very long time for students to learn all the language they might meet in order to read challenging texts such as those proposed by the National Curriculum – and in the meantime, they may well lose interest in the whole endeavour. This then implies that it is important for students to develop effective strategic behaviour to enable them to access more challenging – and thereby, potentially more interesting and engaging – foreign language texts.

Research suggests that strategy instruction can improve foreign language reading comprehension and therefore be an effective use of lesson time (Plonsky, 2011; Ardasheva, Wang, Adesope, & Valentine, 2017). However, to our knowledge, only one study (Macaro and Erler, 2008) has directly evaluated a reading strategy intervention in MFL classrooms in English secondary schools. This reflects the paucity of language learning research conducted in this context more generally. Most research tends to focus on learners of English, who are often adults at higher proficiency levels and with greater extrinsic motivation to learn the language.

Macaro and Erler's (2008) study was an a small scale and exploratory one, in which three classes received explicit strategy instruction in French, integrated into their MFL lessons over part of Years 7 and 8. A comparison group of three classes followed their usual MFL syllabus. The strategy instruction programme, taking around 10 minutes per week, included various strategies designed to help learners identify individual words, use the wider context effectively and monitor their on-going understanding of the text (e.g. 'sound out the word or phrase'; 'guess the meaning of a problem word from surrounding words'; 'read the whole sentence to see if it makes sense'). To support the 'sounding out' strategy, students also received phonics instruction in certain French grapheme-phoneme correspondences (GPC) – that is, the relationship between written forms and the sounds they represent.

The researchers found that the intervention group (who received the strategies instruction) outperformed the comparison group on a challenging reading comprehension task at the end of the intervention, with a large effect size¹. Students in this group were also notably more willing to tackle this text in the first place. Further, there were changes in participants' strategic behaviour, with those in the experimental group becoming more likely to engage actively with the text, and those in the comparison group less likely to do so.

However, the study had a number of limitations, including the following:

- (1) We cannot disentangle the effects of the different components of the intervention. Was it really participants' improved strategic behaviour that made the difference, or were any improvements in their knowledge of French GPC (which were not measured) also important?
- (2) If participants did become better at 'sounding out', this may have led them to recognize more words in the texts, thus unlocking new possibilities for strategic behaviour even without any explicit strategy instruction.
- (3) As part of the instruction programme, students in the intervention group were exposed to a number of challenging texts, whilst those in the comparison group were not. This is because, it was argued, it would have been unfair to give such difficult texts to students without providing support in developing the strategies needed to access them. Whilst this may be so, it nonetheless introduces a potential confounding variable, because it could have been that exposure to more challenging texts in itself (rather than the explicit strategy instruction per se) allowed students to improve their strategic behaviour, as well as their reading outcomes and their willingness to tackle such texts in the first place.
- (4) The researchers themselves were involved in delivering the intervention, not just the usual class teachers. This extra attention from outside visitors may have had a positive effect on students in the intervention classes, perhaps leading them to try harder in the research tasks such as the reading comprehension tests. Researchers also gave individual feedback to learners on their strategy use, which Macaro and Erler (2008) identify as a key element of the intervention's success, which may go beyond what individual teachers may have time to do.
- (5) The sample of schools involved in the study was small, comprising only 6 schools.

Phonics in French

In recent years, phonics teaching has become firmly embedded in the teaching of early reading in English. Perhaps because of this, phonics has begun to spread to MFL classrooms. It has been strongly advocated by people such as Lynn Erler (2004), following her seminal research in this area, and Rachel Hawkes via her popular

¹ An effect size is a standardized measure of the magnitude of a difference between two sets of scores (e.g. the scores of two different groups on the same test). Where we refer to 'effect sizes' in this report, we usually use a measure known as 'Cohen's d'. A conventional rule of thumb says that 0.2 and above can be considered a 'small effect'; 0.5 and above, a 'medium effect'; and 0.8 and above, a 'large effect'.

language pedagogy website². We also know that in the absence of explicit phonics instruction, many Key Stage 3 MFL students are poor at phonological decoding in French (they are error-prone and unconfident when reading aloud) and make little or no progress in this area, even over a period of years (Erler & Macaro, 2012; Woore, 2009, 2016). They do not seem to pick up French decoding incidentally, simply through exposure to the spoken and written language.

Why does being able to decode in a foreign language matter? For beginner readers in their first language, being able to sound out written words (at least the regularly spelled ones) gives access to their spoken forms, which they can then recognize – but this is because they already have a strong grasp of the spoken language at the point of learning to read. Foreign language students, by contrast, often encounter written text at the very beginning of their language learning. Thus, as Grabe and Stoller (2011) argue, “one benefit of developing accurate letter-sound correspondences is lost in most L2 settings; that is, L2 students cannot match a sounded-out word to a word that they know orally because they do not yet know the word orally” (pp. 36-37). On the other hand, Erler and Macaro (2012) did find that a majority of MFL students in their large-scale national survey at Key Stage 3 said that they sounded out French words that they did not recognize when reading – although how successfully, we do not know. Further, evidence is beginning to emerge that accurate phonological decoding is linked to motivation for language learning more broadly (Erler & Macaro, 2012) and that it helps foreign language learners in the all-important task of vocabulary acquisition, both intentional and incidental (e.g. Hamada & Koda, 2008, 2011).

To date, however, very little systematic research has actually addressed the question of whether phonics can be taught successfully in a foreign language context, and if so, what the effects of this might be. An exploratory study based in ten Year 7 MFL classrooms (Woore, 2011) found support for the positive impact of phonics instruction on students’ foreign language decoding. The teaching was delivered by class teachers in 10-minute ‘segments’ each lesson for around 30 lessons. The intervention group (who received this instruction) made significantly more progress than the comparison group in French decoding. This study did not, however, assess the impact of the phonics instruction on students’ ability to comprehend written French (or on other aspects of their language learning).

2. Aims of our study

It is imperative that beginner MFL learners in Year 7 quickly develop the foundation literacy skills they need to support their subsequent progress in foreign language reading, as they move through Key Stage 3 and towards KS4. Their MFL teaching must also help them feel they are making real progress, and thus sustain the high levels of motivation which they often bring with them from primary school. With these goals in mind, our study aimed to strengthen the evidence base on which teachers can draw when teaching reading in Year 7 MFL. We focussed on French, since this is by far the most widely-taught MFL in England (Board & Tinsley, 2015), yet its spelling-to-sound mappings are challenging for English students.

Specifically, we wished to build on the exploratory studies based in English MFL classrooms, mentioned above, whilst also addressing some of their limitations. In particular, in relation to reading comprehension outcomes, we sought to tease out the individual contributions (if any) of a phonics-only programme and a strategies-only programme, as opposed to the combined strategies-and-phonics programme implemented by Macaro and Erler (2008). We also wished to find out whether any improvements in phonological decoding which might result from French phonics instruction would have any impact on reading comprehension or other aspects of language learning. Finally, we were interested to know whether any benefits of explicit strategy instruction was actually due to that instruction itself, or whether the same outcomes could be achieved simply through giving students practice in tackling more challenging texts.

Our research questions were:

1. To what extent does a programme of explicit instruction in (a) reading strategies or (b) phonics lead to improvements in French reading proficiency amongst Year 7 MFL students, as compared to a programme of reading challenging texts?

² www.rachelhawkes.com/Resources/Phonics/Phonics.php

2. To what extent do any of the three programmes of instruction lead to improvements in: (a) students' French phonological decoding proficiency; (b) their strategic behaviour when reading in French; (c) their vocabulary knowledge; and (d) their motivation for learning the language?
3. Is one of the programmes of instruction more effective than the others on any of the above measures?
4. What are teachers' and students' views on the three programmes of instruction?

3. Methods

Overview

We conducted an experimental study, and more specifically a Randomized Control Trial (RCT). This is a type of study adapted from medical research, in which participants are randomly allocated, in equal numbers, to each of the experimental groups. This study design is often described as the 'gold standard' for establishing causal relationships such as the effects of particular interventions (Moore, Graham & Diamond, 2003). There have been calls for greater use of RCTs to evaluate educational interventions (e.g. Hutchison and Styles, 2010). However, such studies do also have limitations – for example, by focussing only on a limited number of outcomes, or by looking only at the average effects on a group as a whole and losing sight of variations between individual participants.

In our experiment, there were three groups:

1. Strategies: explicit strategy instruction, supported by the use of challenging texts;
2. Phonics: explicit phonics instruction, again supported by the use of the same challenging texts;
3. Texts: use of the same texts as above, but with no explicit instruction in either strategies or phonics.

We would ideally have liked to have two additional groups: one receiving instruction in both strategies and phonics, and one 'true control' condition following their usual curriculum, without strategies / phonics instruction and without access to the challenging texts. However, this was not feasible within the time and resources available to us. Further, these additional two groups were the ones included in Macaro and Erler's (2008) study. We therefore conceptualized our study as building on this previous study, with their two groups acting as points of comparison for our own findings.

Outcome measures

We tested students' outcomes immediately before the teaching (November-December 2016), immediately after the teaching (June-July 2017) and again some time later (November-December 2017) to see how durable any effects of the interventions are. We collected almost all data³ 'blind to condition': in other words, the person administering or marking the tests did not know which group the participants concerned were in, in order to guard against any bias.

The main outcome variables that we measured, and the instruments used to measure them, are listed below.

1. Reading comprehension, measured using a mixture of (a) translation tasks into English and (b) comprehension exercises, in which students answered some questions in English about the 'main ideas' in short French texts.
2. Phonological decoding, measured using a 'Sound-Alike Task' (SALT). This is a pen-and-paper test which we developed especially for the study, in which participants were presented with sets of three French nonsense words and had to decide which (if any) sounded the same.
3. Vocabulary knowledge. We tested the breadth of students' vocabulary knowledge (i.e. how many French words they know) using a simple and widely-used task called X_Lex (Meara & Milton, 2003).
4. Strategic behaviour, measured using a self-report questionnaire adapted from Macaro and Erler (2008).

³ Almost all data collection was carried out blind to condition, with one or two minor exceptions, which were unavoidable due to practical constraints. All marking was done blind to condition, with no exceptions.

5. Self-efficacy and motivation for both reading French and for learning French more widely, measured using self-report questionnaire items adapted from Macaro and Erler (2008) and Graham and Macaro (2008).

We also collected students' and teachers' views on the pedagogical interventions via questionnaires (completed by all participants) and interviews (completed by a small sub-sample of students and teachers). Additionally, we administered a 'background information questionnaire' to gather some contextual information on our participants (e.g. whether they spoke other languages at home; whether they had any French-speaking parents or had lived in France). All of the instruments used in our study are being made available to download from our website (pdcinmfl.com), except for the X_Lex (for which see <https://www.lexutor.ca/tests/>).

Finally, we accessed students' existing academic records (such as their CAT scores, reading ages and primary school SAT scores) as held by their schools, where we were able to do so. This was to allow us to use students' prior attainment as a contextual variable in our analysis, because it might be expected to influence their scores on our tests. Unfortunately, we were unable to obtain prior attainment data from 8 schools (due either to administrative difficulties or to the schools' concerns over sharing sensitive data). Others sent data but with gaps (missing values) for individual students. Further, not all of the remaining 28 schools held the same academic data: for example, some had CAT but not SAT scores, and others vice versa. To achieve the largest possible usable 'prior attainment' dataset, we therefore took both of these variables (CATs / SATs) and 'standardized' them, in order to place them onto a common scale and allow us to make comparisons between them⁴. Using this method, we obtained prior attainment data for 693 participants.

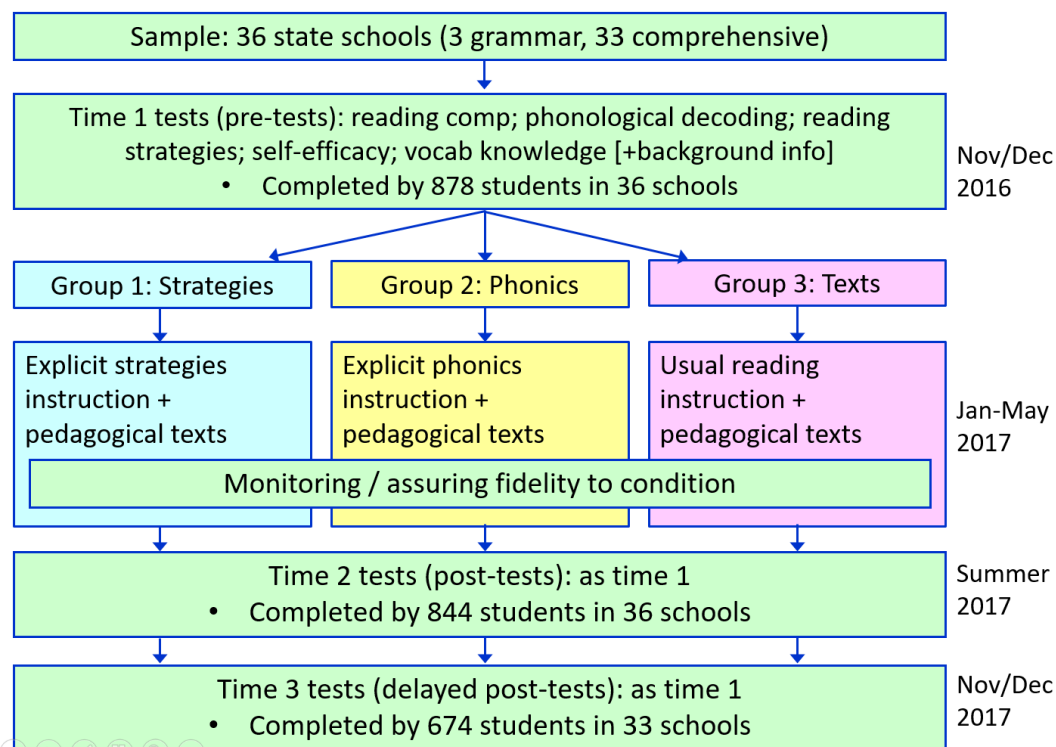
Allocation of classes to experimental groups

Our study is a 'cluster trial', in which whole classes of students (each with a different teacher, and each in a different school) are allocated to the different groups. We did this using a process called 'minimization' in order to avoid some of the drawbacks associated with simple randomization, such as imbalances between the conditions which might occur by chance (Torgerson & Torgerson, 2013). For example, our study included three grammar schools, and it would not have been good if, through a simple randomization process, two or three of these had ended up in the same group. Minimization works by randomly allocating the first school to one of the three groups, and then subsequently allocating schools to groups in such a way as to create groups which are as well-matched as possible. In our case, these variables were: (a) school type (grammar or non-selective); and the percentages of students in each school who: (b) were eligible for Free School Meals; (c) were recorded as having English as an Additional Language; (d) achieved five or more A* to C grades at GCSE, including English and Maths.

A diagrammatic overview of the FLEUR study design can be found in Figure 2. (A more detailed flow chart, following the CONSORT guidelines (<http://www.consort-statement.org/>) and showing more detailed sample sizes at each stage of the project, can be found in Appendix 1).

⁴ To standardize the prior attainment data, we converted participants' average CAT scores – and where these were unavailable, their average SAT scores – into 'z scores'. To do this, we subtracted the sample's mean value for each of these variables from each participant's score on that variable, and divided the resulting number by the standard deviation of the sample for that variable. The z score tells us how far a given individual's score is above or below the group mean. For example, a z score of 0 is identical to the group mean; a score of +1 is 1 standard deviation above the group mean.

Figure 2: overview of the FLEUR study design



Sample: participating schools

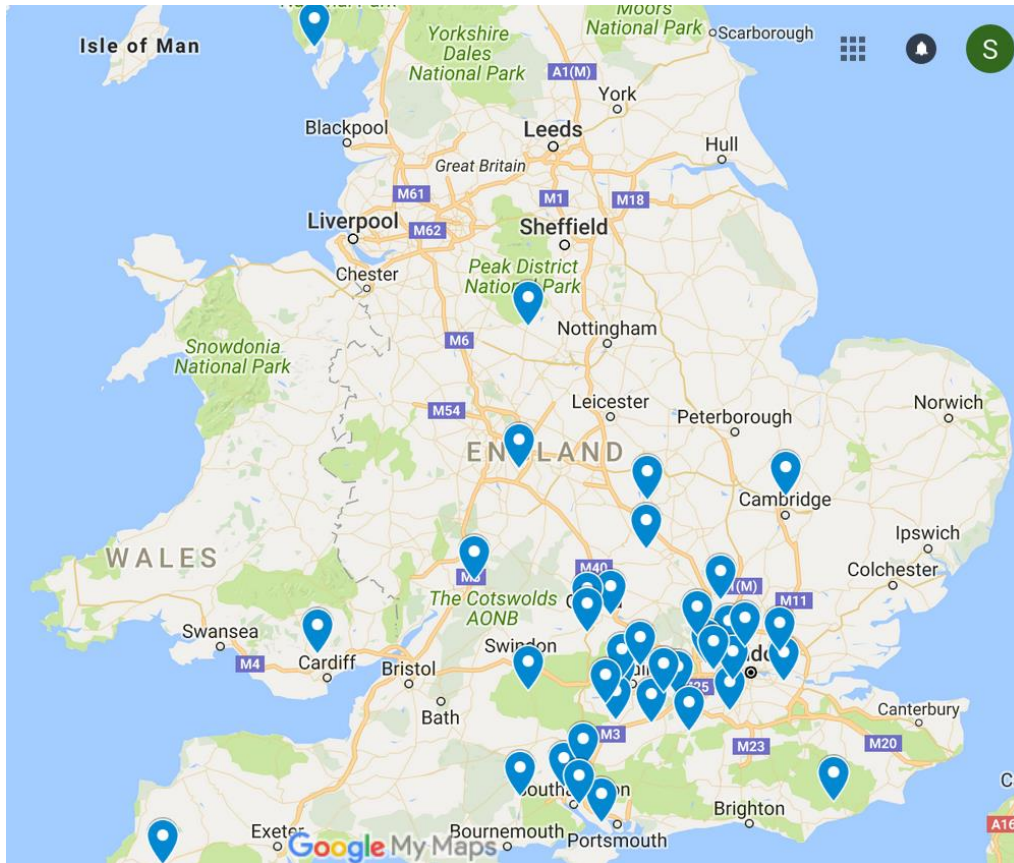
We recruited 36 state schools to take part in our study, including 3 grammar schools and 33 non-selective. Our sample included a diverse range of schools spread over a wide geographical area (Figure 3). Their characteristics are summarized in Table 1, based on publicly available data for the 2014-15 school year, the latest for which data was available at the point of allocating schools to the experimental groups. Note that the values are heavily affected by the inclusion of the three grammar schools in our sample, which are high attaining schools with low numbers of students eligible for Free School Meals or receiving SEN support.

Table 1: characteristics of schools in the FLEUR study

	Minimum in our sample	Maximum in our sample	Average for our sample	National average ⁵
(a) % students achieving five or more A* to C grades at GCSE (including English and Maths)	40	100	68.8	53.8
(b) % students eligible for Free School Meals	1.6	43.5	17.4	29.4
(c) % students recorded as having English as an Additional Language	0.2	97.8	13.4	15.0
(d) Percentage of students with support for Special Educational Needs	1.4	26.9	11.2	unknown

⁵ The national average figures were taken from official government statistics, available online at <https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2017> and <https://www.gov.uk/school-performance-tables>

Figure 3: approximate locations of the FLEUR schools



In total, almost 900 students took part in our project (though not all of these participants completed all of the data collection: see also 'Attrition' below). The sample was spread roughly (but not perfectly) evenly between the three experimental groups. The average number of students who participated per class was 25.

Interventions: the programmes of instruction

The pedagogical interventions in our study lasted for a period of 16 weeks, beginning in January 2017 and finishing in Summer 2017. The aim was for teachers to spend 20-30 minutes per week on the intervention teaching, on average, during this period. As noted above, the content, nature and format of the interventions were designed collaboratively with the participating teachers in each group. All of the materials (including resources and lesson plans) are being made freely available through our pdcinmfl.com website. A brief description of each intervention is provided below. Note that teachers also agreed to refrain from giving explicit instruction pertaining to the other experimental groups for the duration of the intervention (e.g. the strategies group taught no explicit phonics, even if they would usually have done so).

Strategies

The strategies group began with some tasks designed to raise their awareness of the kinds of strategies they could use to help them understand challenging texts. For example, one of these tasks asked them to read a short English text with some nonsense words embedded in it, and then to work out the meanings of the nonsense words. The strategies introduced were then given to students as a checklist of 8 strategies (available in both English and French) which were revisited throughout the intervention.

Subsequently, students worked with a series of eight 'pedagogical texts' that we had created. These covered topics related to francophone culture and were intended to be linguistically challenging (but not to the point of being inaccessible) for Key Stage 3 MFL students. We hoped that students would feel that they were genuinely learning something new from reading the texts (rather than seeing foreign language reading simply in terms of

'language practice') and thus find them engaging. The texts also had links to related video clips on Youtube, designed to pique students' interest or further develop their knowledge of francophone culture.

Figure 4: sample Powerpoint slide from the strategies intervention

L'éléphant géant et le dragon-cheval de Nantes

I have:

1. Made a real effort to persevere and understand the text
2. Read on to look for the meaning of words I don't know
3. Looked at the co-text (the words before or after the difficult word)
4. Used logic to work out a word's meaning
5. Used my common sense, my own experience, my 'world knowledge'
6. Used my knowledge of English words
7. Used other clues such as pictures or a title
8. Checked that my interpretations make sense

Nantes est une ville française de 300 000 habitants

qui est située sur les rives de la Loire.




Most sessions began with some explicit modelling of strategy use by the teacher ('thinking aloud') using a short portion of the text (e.g. the first paragraph) to show how the 8 strategies could be used to help them understand (see sample Powerpoint slide in Figure 4, which covers the first sentence in one of the texts). Students would then tackle additional paragraphs (in groups, pairs or individually, as the teacher deemed appropriate), and were encouraged to continue to use a range of reading strategies whilst doing so. After reading, they went through the general meaning of the text with the teacher, as well as reflecting together on the strategies they had used and how effective they had been. Students were asked to set themselves goals to achieve in their strategy use and gave each other feedback on their progress towards these goals. However, unlike in Macaro and Erler (2008), students did not receive individualized written feedback on their strategy use, because participating teachers felt that this would not be possible given time constraints and existing workload. This is a limitation of our study in terms of its attempt to tease apart the effects of the strategy and phonics instruction as implemented in Macaro and Erler's study.

Phonics

The Phonics intervention again began with some awareness raising tasks, designed to help students notice and reflect on some of the differences between French and English grapheme-phoneme correspondences – for example, the fact that the word *nation* is spelt identically in the two languages, but some of the letters are pronounced quite differently. Subsequently, a set of specific grapheme-phoneme correspondences were introduced each week, using short 'exemplar' words to practise and embed the correct pronunciations (see sample Powerpoint slides in Figure 5). Students were also asked to reflect on any other words they knew containing these 'target' GPC, and then to practise reading aloud some sentences containing them – first under their teacher's guidance, and then individually or in pairs / small groups.

Finally, every few lessons, students worked with the same pedagogical texts as those used in the Strategies intervention (to ensure equal exposure to challenging texts across the groups). In addition to facilitating the deployment of certain strategies, each text had also been designed to exemplify particular GPC: for example, the text on *L'éléphant géant et le dragon-cheval de Nantes* includes many words containing the grapheme <an>. Teachers did not, however, encourage or support students in the use of strategies to comprehend the texts. Instead, students answered a series of basic questions designed to be easily accessible, for example by

recognizing numerals or clear cognates. We also provided some more open-ended, reflective questions intended to really make students think and engage with the cultural content: for example, was Philippe Petit’s famous tightrope walk between the twin towers a piece of art, as he claimed?

Figure 5: sample slides from the Phonics intervention (produced using Microsoft Powerpoint)



Figure 6: an example of one of the FLEUR ‘pedagogical texts’

L’éléphant géant et le dragon-cheval de Nantes

Nantes est une ville française de 300 000 habitants qui est située sur les rives de la Loire.

Dans le passé, c’était un port important avec accès à l’océan atlantique. À l’époque, Nantes était un grand centre industriel avec de nombreux chantiers navals. Mais Nantes a beaucoup changé. C’est maintenant une grande destination touristique.




L’une des attractions touristiques les plus bizarres, c’est « Les Machines de l’Île », un projet artistique. Ça se trouve dans le quartier du vieux port. On y trouve des animaux mécaniques, tels qu’un éléphant gigantesque qui marche dans les rues entouré de touristes et vacanciers. Il mesure 12 mètres de hauteur et pèse 48 tonnes. Il peut transporter 50 passagers en tout.




Depuis 2014, il y a un nouvel animal géant. Mi-cheval, mi-dragon, cette créature mythique s’inspire d’une légende chinoise. Elle s’appelle « Long Ma » : ça signifie « dragon cheval » en mandarin. Contrôlé par un chauffeur, il fait des promenades dans la ville de Nantes. Avec les flammes et la fumée qui émanent de sa bouche, c’est un spectacle magnifique – mais aussi un peu menaçant et même effrayant!

C’est intéressant ? Tu peux regarder un clip vidéo ici : www.youtube.com/watch?v=pHnMfKyh340

Texts

As already mentioned, the Texts group received no explicit instruction in either strategies or phonics. They did, however, work with the same pedagogical texts as those used in the other two interventions. They answered the simple, factual questions and open-ended, reflective questions (as described above), and watched the Youtube clips in class. An example of one of these texts is shown in Figure 6, but all are being made freely available through our website.

Responsive teaching and 'fidelity to condition'

There are a number of issues which make experimental trials in education much more complex than in the medical field. One is that, when trialling a new medicine, exactly the same drug in exactly the same dose is administered to all participants in the 'intervention' group. In our study, however, the pedagogical interventions were delivered to whole classes by their respective teachers. This introduces scope for variation between the different teachers, school contexts and indeed groups of students involved (with the latter interacting with the planned teaching in distinctive and often unpredictable ways). We took two steps in relation to this issue:

1. We acknowledged that some variation in the implementation of the interventions was inevitable, and indeed in some senses desirable. This is because all classes are different; effective teaching will always respond to the needs of particular students on particular occasions. Therefore, we did not attempt to create a uniform, 'off-the-shelf' intervention which we would impose on teachers and ask them to implement uniformly. Rather, we worked with participating teachers at the start of the project to agree on the nature of the interventions (what material would be covered, in what amount of time and what the broad pedagogical approaches would be); we then created a series of lesson outlines which we gave to all teachers, but which they were able to adapt to suit the needs of their own particular class. We met with teachers periodically in order to discuss how the interventions were going and to determine whether any changes or 'tweaks' were necessary. We also offered support to teachers in all three groups via email and telephone.
2. We monitored the ways in which the interventions were actually delivered in the different schools, by asking teachers to send us records of the nature and extent of their French reading instruction with participating classes, and by observing selected lessons. This enabled us to assess 'fidelity to condition': how 'faithfully' the actual teaching that students received matched what had been intended when designing the study.

Based on the information we collected, we assigned each school a fidelity to condition 'band' using a three-point scale:

1. Excellent. All or almost all of the planned material was covered, including the pedagogical texts. A large amount of time was spent on target teaching, with no (or hardly any) non-target teaching.
2. Medium. All or almost all of the planned explicit instruction was covered, but there may also have been some non-target teaching, and/or some pedagogical texts may have been omitted (with between 4 and 6 being completed).
3. Poor. Very little or none of the planned material was covered. Hardly any time was spent on the target teaching, and/or there was equally as much non-target teaching as there was target teaching, and/or fewer than half of the pedagogical texts were covered.

We judged 25 of our 36 schools to have had excellent fidelity to condition (band 1), and 11 medium (band 2). The most frequent reason for slipping into band 2 was not covering at least 7 of the pedagogical texts. In turn, this was generally due to unavoidable factors such as teacher illness or absence. We were pleased that no schools were judged to have poor fidelity to condition (band 3).

We found no evidence of any systematic relationship between group and fidelity to condition. In other words, the schools in bands 1 and 2 were evenly distributed across the three experimental groups. The mean number of texts covered by each group were 7.3 (strategies), 7.4 (phonics) and 6.9 (texts).

Overall, we are satisfied that our three interventions were implemented with sufficient fidelity to enable valid comparisons to be made between them.

Attrition

A large majority of the students involved in the pre-tests (henceforth referred to as 'time 1') also completed the post-tests (time 2), although some students were lost due to opting out or absence on the day of data collection. The total number of students who completed the reading comprehension tests was 878 at time 1; of these, 804 also completed the tests at time 2.

There was, however, significant attrition at delayed post-test (time 3), and to a much larger extent than we had anticipated. We lost three entire schools (one in each experimental group) as well as a number of students in other schools. In the overwhelming majority of cases, this was for unavoidable logistical reasons: for example, classes had been 'settled' by attainment or otherwise reorganized, or had changed from French to another modern language; or the participating teacher had changed schools, and the new class teacher was unable to support the study. The total number of students who had completed the reading comprehension tests at times 1 and 2, and also completed them at time 3, was 674.

As a result of this attrition, we have less confidence in any findings involving time 3. We therefore consider our time 1 – time 2 comparisons to provide the most robust findings from our study.

More detailed information on the sample sizes at each stage of the project can be found in the CONSORT-style flow chart in Appendix 1.

Analysis

For all our quantitative outcome variables, we used statistical tests to compare the scores obtained by the three groups at each time point, and to compare the scores within each group across the different time points. In these comparisons, we looked for 'statistically significant' differences: that is, differences that are highly unlikely to have arisen simply through chance variations in the data. So, for example, if we found no significant differences between our groups' reading comprehension scores at time 1, but then found that one group had a statistically significant advantage (higher score) at time 2, then we could confidently attribute this difference to the effects of the intervention. In the output of statistical tests, according to convention, statistical significance is indicated by a 'p' value lower than .05 (i.e. there is only a 5% chance that the given result occurred by chance). Where we report 'statistically significant' findings below, we therefore mean that $p < .05$ ⁶.

In many of our between-group comparisons, we also used a statistical approach known as 'multilevel modelling', with two levels: individual participants at level 1 and schools at level 2. This approach tests for statistically significant differences between the groups' scores, whilst taking into account the fact that there may be a 'school-level' effect on students' outcomes. For example, students in the same school may be similar in some respects, because they are from a similar demographic and the same area. Further, in our study, there was one class per school, so all the students in a given school were also in the same French class, with the same classmates and the same teacher. Their outcomes may therefore be influenced by their teacher's overall characteristics and instructional approach beyond the programmes of instruction implemented as part of our study. (This is known as a 'teacher effect'). In many analyses, we also controlled for participants' prior academic attainment (where this data was available), since this might also be expected to influence their outcomes.

In the interests of readability, we have generally not included full details of the statistical tests in the main text of our report. However, where we feel such details may be useful (e.g. so that interested readers can see which variables were included in our statistical models), we have provided them in footnotes or appendices.

⁶ In some cases, a more rigorous statistical threshold was used to take account of the fact that we were making multiple statistical comparisons on the same dataset.

Ethical issues

Our study adopted followed all procedures expected for research of this kind, to ensure the highest ethical standards. We followed the ethical guidelines for educational research produced by the British Educational Research Association (BERA, 2011): for example, voluntary informed consent for all participants; secure data storage; anonymization of all data.

The FLEUR study is registered with ISRCTN, an internationally recognized registry of clinical trials which “supports transparency in clinical research, helps reduce selective reporting of results and ensures an unbiased and complete evidence base” (<http://www.isrctn.com/>). The unique reference number for our study is ISRCTN11400585.

3. Findings

Reading comprehension

The first main outcome that we wished to test was reading comprehension. We wanted to know whether any of the programmes of instruction (Strategies, Phonics, Pedagogical Texts) led to greater improvements in reading comprehension than the others.

Our primary analysis is based on ‘intention to treat’. In other words, we included in our analysis all the schools who were originally allocated to each condition, irrespective of any differences in the teaching actually received by students in each group (for example, if the programmes of instruction were implemented more ‘faithfully’ in some schools than others). Analysis based on intention to treat is seen as a fairer test of the programmes of instruction, because it reflects the fact that in the ‘real world’ beyond formal experiments like this, variations in the implementation of a given instructional programme are also likely to occur. Therefore, removing from the analysis any schools where (for whatever reason) the programmes of instruction are less faithfully implemented might give a biased view of the effects of the programme.

Further, because of the attrition we experienced from our sample at time 3 (the delayed post-test, in December of Year 8), we base our main analysis on the effects of the interventions between times 1 and 2 (immediately before and after the programmes of instruction were implemented). This also reflects the fact that any effects of the interventions may be expected to get diluted or ‘washed out’ after the long Summer holiday, followed by a term of teaching under different instructional approaches (and, in many cases, with a different teacher).

Reading comprehension tests at times 1 and 2: students’ progress

The reading comprehension tests at times 1 and 2 both comprised: (a) two short translation tasks and (b) two short ‘main ideas’ tasks. One of the translation tasks, plus both main ideas tasks, were common across the two time points. It is on these common tasks that we base our main analysis, because this provides direct insight into students’ progress across the two time points (since we know that these tasks are of identical difficulty at time 1 and time 2). However, we also report additional analyses based on participants’ overall scores (i.e. using all four reading tasks at each time point).

We realize that some people might argue that this direct repetition of tasks could cause a ‘practice effect’ – i.e. participants might get better just because they have done the same tasks twice. However, we do not believe there is much risk of this, given that there was a six month delay between the two tests, and in any case students did not receive any feedback on their first attempt.

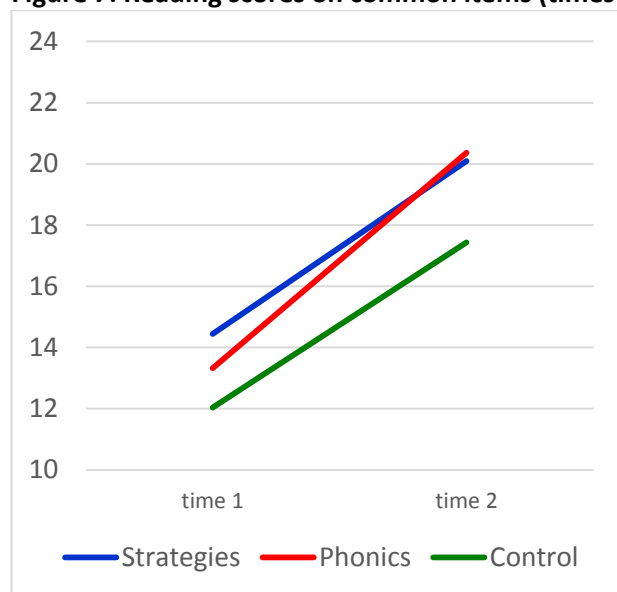
On average, on the common tasks across times 1 and 2, the sample as a whole increased their score by 6 marks (out of a total of 45 marks). Further, all three of the experimental groups increased their scores, and statistical tests (repeated measures ANOVA) showed that these increases were statistically significant in each case, with

large effect sizes⁷ (Table 3, Figure 7). In other words, the time 1-time 2 score difference for each group is both large and highly unlikely to have occurred by chance. It may therefore be that working with our ‘pedagogical texts’ was in itself beneficial for developing students’ French reading proficiency, irrespective of whether or not they additionally received explicit instruction in phonics or strategies.

Table 3: mean scores on common reading comprehension items at times 1 and 2

	N	time 1	SD	time 2	SD	mean change	F	p	η_p^2
Whole sample	804	13.26	8.76	19.34	9.80	6.1	676.28	<.001	.46
Strategies	255	14.4	8.04	20.09	9.25	5.7	226.50	<.001	.47
Phonics	292	13.3	9.33	20.36	10.53	7.1	310.18	<.001	.52
Texts	257	12.04	8.64	17.43	9.20	5.4	158.67	<.001	.38

Figure 7: Reading scores on common items (times 1 and 2)



We can also see from Figure 7 that the line for the Phonics group is steeper, suggesting that on average they made more progress (an average increase of 7 points) than the Strategies and Texts groups (an average increase of around 5.5 points). It is therefore possible that explicit Phonics instruction had a positive impact on reading outcomes over and above simple exposure to the pedagogical texts.

However, when a multilevel statistical test was conducted to take account of possible school and teacher effects, we found no statistically significant difference between the groups in terms of their reading comprehension scores (on the common items) at time 2, after controlling for their time 1 scores. (We did this because some of the variation in scores at time 2 is predictable based on how well individual students did at time 1). In other words, we cannot say with any confidence that any of the three programmes of instruction (phonics, strategies or texts) led to greater improvements in participants’ reading comprehension than the others. For details of the multilevel regression model that we conducted, please see appendix 2.1.

We ran an additional analysis using the same outcome data (time 2 scores on items common across times 1 and 2), but including a measure of students’ prior attainment (z scores based on CAT or SAT data) as a covariate, to control for the likely influence of this variable on students’ outcomes. There remained no significant difference between the groups in order to their time 2 reading scores, after controlling for their time 1 reading scores. Again, details of these tests can be found in appendix 2.1.

⁷ See footnote 1 for an explanation of ‘effect sizes’. Here we use an effect size measure called ‘partial eta squared’ (η_p^2). According to conventional rules of thumb (see <http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize>), a partial eta squared value of 0.02 indicates a small effect, 0.13 a medium effect and 0.26 a large effect.

We also repeated the same multilevel analyses as described above to examine participants' overall reading comprehension scores at each time point (based on all four tasks at each time point: the two translations and two main ideas passages). Again, we found no statistically significant differences between the groups' scores either at time 1 or at time 2 (after controlling statistically for their time 1 scores). Again, therefore, this analysis provided no evidence that any of the programmes of instruction was more effective in improving reading outcomes than the others. Details of these analyses are provided in appendix 2.2.

Reading comprehension tests at times 2 and 3

The time 3 test comprised two translations, one of which was identical to one of the time 2 translations. As noted above, there was considerable attrition at this point in the project, reducing our confidence in our findings. Statistical tests showed no significant differences between the three groups at either time 2 or time 3.

Again, for more direct insight into participants' progress, we compared their scores on the common translation task between times 2 and 3. Overall, we found that all groups increased their score by about 1 mark, out of a possible total of 35 marks. In other words, there was a levelling off of progress compared to the larger increases found between times 1 and 2. Viewed from another perspective, it could be seen as promising that progress levelled off rather than dropping, given that considerable time elapsed between times 2 and 3, and that this period including the long Summer holidays, when students received no French instruction in school.

Willingness to 'have a go'

Finally, we were interested to see whether (as had been found in Macaro and Erler, 2008) there was any evidence that students following a particular programme of instruction (strategies, phonics, texts) became more willing than those in the other groups to tackle the kinds of challenging texts used in our reading comprehension tests. We found no evidence of this in our study. At each time point, there were roughly equal (and relatively low) numbers of students in each group who, despite being present for the data collection, left their script entirely blank for one or both translation tasks, indicating that they had not even attempted to translate it.

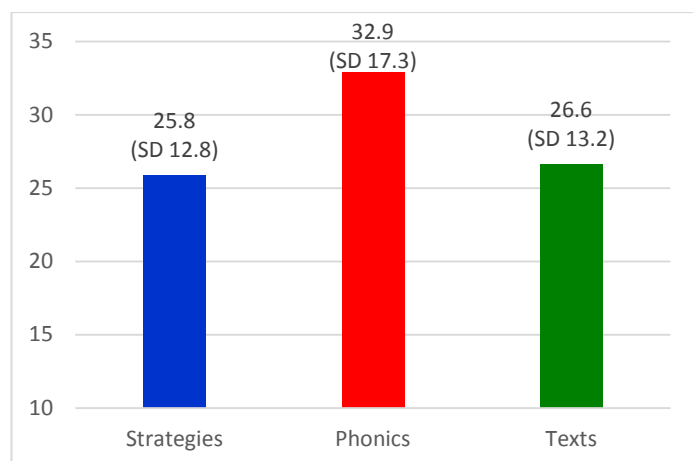
Phonological decoding

We measured phonological decoding using a pen-and-paper test that we had developed (the Sound-Alike Task, or SALT). Two different versions were used, one at time 1 (comprising 14 items) and the other (19 items) being the same across times 2 and 3.

We found no statistically significant differences between the groups' decoding scores at times 1 or 3. Analysis of time 2 decoding scores using a multilevel model, in which we controlled for time 1 decoding scores, also found no evidence of significant differences between the groups (Appendix A3, Model 8). However, there was an advantage for the Phonics group over the Texts group which was on the cusp of statistical significance ($p=.053$, where a result of $p < .05$ is conventionally considered significant). When a further model was conducted which controlled for students' prior attainment, there remained no significant advantage for being in the Strategies group over the Texts group, but the advantage for the Phonics group over the Texts group was statistically significant, $p = .038$ (Appendix A3, Model 9). On average, whilst participants in the Strategies and Texts groups performed similarly on decoding task at time 2, the Phonics group scored 6-7 percentage points higher (Figure 8).

We would note at this point that it is likely that our SALT was a less sensitive measure of phonological decoding than a Reading Aloud Task would have been. The latter, however, is extremely time consuming to administer and mark, and so was not feasible to use with our large sample of students. Had it been feasible to do so, however, we consider it likely that an even clearer advantage would have been found for the Phonics group over the other two groups.

Figure 8: Mean percentage score on the time 2 decoding test, subdivided by group



We included in the SALT some items which were common to the different versions of the test used at times 1 and 2. This allowed us to use a statistical approach (Rasch modelling) to use these items as ‘anchors’ and thus put the two test versions onto a common scale of difficulty. In turn, this provided insight into students’ progress in phonological decoding over time. We found that all three groups showed statistically significant score increases between times 1 and 2⁸. However, the magnitude of the improvement was small for the Strategies group (with an effect size of $\eta_p^2=0.02$) and for the Texts group ($\eta_p^2 = 0.07$). For the Phonics group, the effect size was larger, almost reaching the conventional threshold for a medium-sized effect ($\eta_p^2 = 0.12$).

At times 2 and 3, the same version of the test was used, allowing us to make direct comparisons of the scores at each time point. All three groups increased their scores slightly (Strategies group: + 1.6 points out of 19; Phonics group + 0.5; Texts group + 1.4). However, these improvements were not statistically significant in any of the groups.

Vocabulary knowledge

We administered the same vocabulary test at all three time points, again allowing us to make direct comparisons of students’ scores. The test was a 48-item version of the French X_Lex test, as used by Milton (2006), and adapted for our study in consultation with Professor Milton. The test, which is very quick and easy to administer, provides an estimate of the breadth of participants’ French vocabulary knowledge (i.e. how many words they know) out of the 2,000 most frequent French words. It does this by asking them to indicate whether they know a random selection of real, high frequency French words. A number of nonsense French words are also included in the test, to allow a correction to be applied for guesswork.

One problem with this test is that it relies on students honestly and reliably reporting which words they know. For students (like ours) with very low levels of language knowledge, there is a risk that they may be prone to high levels of guesswork, which can make the test unstable. For example, it can paradoxically result in negative estimates of their vocabulary knowledge (e.g. it might estimate that they know minus 100 French words, which is clearly impossible). Where we found evidence of excessive guesswork like this, we removed those individuals from the analysis⁹. This applied to about 11% of our sample at times 1 and 2, and about 10% at time 3.

⁸ Repeated measures ANOVA using ‘person location’ values (derived from the Rasch model) at each time point. Strategies group: time 1 mean = -0.80 (SD 1.16), time 2 mean = -0.65 (SD 0.79), $F(1, 246) = 5.20$, $p = .013$, $\eta_p^2 = 0.03$; Phonics group: time 1 mean = -0.70 (SD 1.21), time 2 mean = -0.28 (SD 1.03), $F(1, 283) = 37.46$, $p < .001$, $\eta_p^2 = 0.12$; Texts group: time 1 mean = -0.95 (SD 1.14), time 2 mean = -0.64 (SD 0.84), $F(1, 248) = 19.81$, $p < .001$, $\eta_p^2 = 0.07$. Conventionally, a partial eta squared (η_p^2) value of 0.02 indicates a small effect, 0.13 a medium effect and 0.26 a large effect.

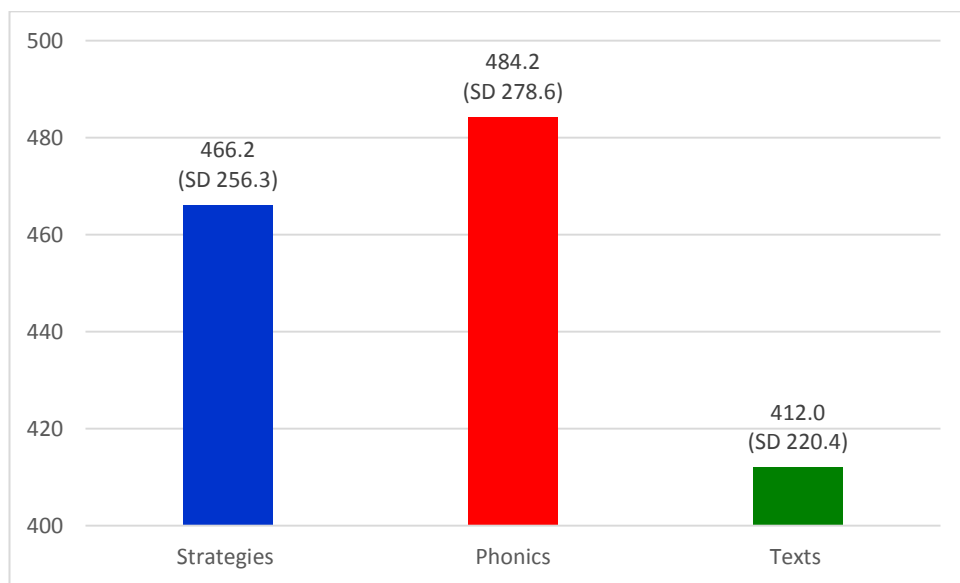
⁹ Following discussion with Professor Milton, we removed any students with a ‘false alarm rate’ (where they incorrectly thought that pseudowords were real words) of >20%, i.e. 3 or more false alarms; and in the few instances where negative scores remained, we corrected these to 0 to indicate that the students concerned had demonstrated no measurable knowledge on the test.

Unfortunately, a consequence of this approach is a reduction in the sample size on which the analysis can be based (as shown in the column labelled 'N' in Table 4).

On a descriptive level, the Phonics group scored highest on the time 2 vocabulary test, followed by the Strategies group and then the Texts group (Figure 9). In order to test for statistically significant differences between the groups, we again conducted multilevel models to take account of school effects, whilst also controlling for participants' time 1 vocabulary scores (i.e. we wanted to focus on only that portion of the variation in time 2 scores which was not predictable based on the time 1 scores). These analyses found no significant differences between the three groups' vocabulary scores at time 1. At time 2, however, a significant difference was found, even after controlling for each group's time 1 scores (Appendix 2.4, Model 11). Specifically, the Phonics group recorded significantly higher vocabulary knowledge at time 2 compared to the Texts group. In other words, this analysis found evidence that the Phonics intervention led to greater vocabulary growth than the Texts-only intervention. When we conducted a further model in which we additionally controlled for prior attainment (which was also a significant predictor of time 2 vocabulary scores), we found that both the Strategies and the Phonics group performed significantly better than the Texts group (Appendix 2.4, Model 12).

We would note at this point that the Phonics group had been exposed to additional vocabulary as part of their phonics instruction, through the Powerpoint slides used to introduce and exemplify particular GPC. Inadvertently, one of these words was also included in the X_Lex test. However, as this is only one word out of the 48 items in the test, we feel that this can be discounted as a confounding variable.

Figure 9: Vocabulary score (out of 2000) at time 2, subdivided by group.



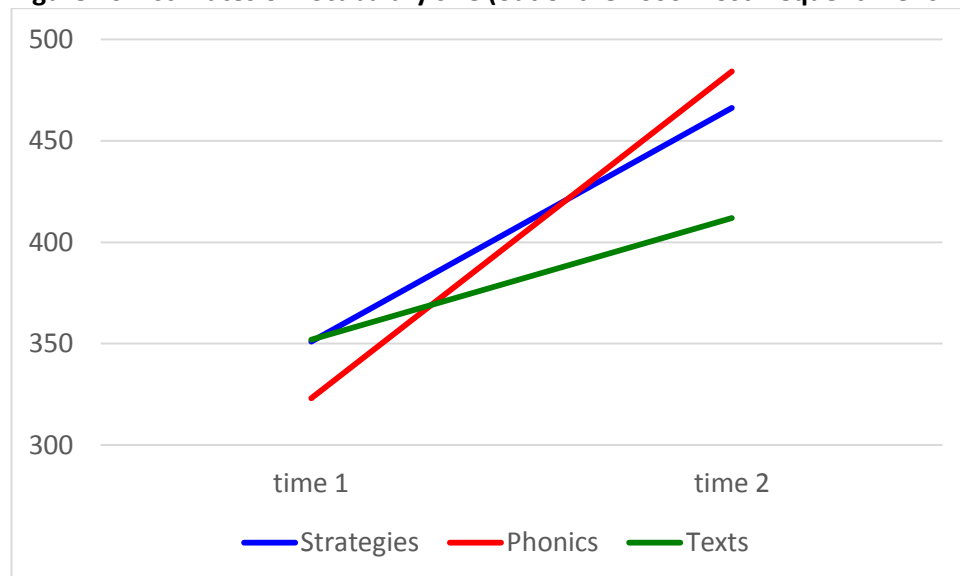
At time 3, analysis using a multilevel model again found no significant differences between the groups, suggesting that some of the differences observed at time 2 had been 'washed out' six months after the end of the interventions (and after returning to their usual curricula).

Looking at each group's scores across time (Table 4, Figure 10), we found a significant increase in recorded vocabulary knowledge for all three groups between times 1 and 2, with a small effect size for the Texts group (in which participants' estimated vocabulary knowledge grew by 60 words on average), a small to medium effect size for the Strategies group (+116 words) and a medium to large effect size for the Phonics group (+160 words).

Table 4: Mean vocabulary size estimates (out of the 2000 most frequent French words) at times 1 and 2

	N	time 1	SD	time 2	SD	mean change	F	p	η_p^2
Whole sample	647	341.0	224.6	455.4	255.6	114.4	96.3	<.001	0.13
Strategies	201	351.0	283.1	466.2	256.3	115.2	27.61	<.001	0.12
Phonics	238	323.1	211.3	484.2	278.6	161.1	66.1	<.001	0.22
Texts	208	351.9	225.9	412.0	220.4	60.1	11.0	.001	0.05

Figure 10: Estimates of vocabulary size (out of the 2000 most frequent French words) at times 1 and 2



Between times 2 and 3, there was a significant change in estimated vocabulary knowledge for the Texts group only, albeit with a small effect size ($p = .002$, $\eta_p^2 = 0.06$): on average, participants in this group increased their vocabulary knowledge by approximate 60 words. (The other two groups registered slight decreases in vocabulary knowledge (Strategies group: -40 words; Phonics: -16 words), but these were non-significant). As in previous analyses of time 3 scores, however, we should bear in mind the reduced sample size at this time point.

Self-efficacy for French reading

At each time point, we assessed students' self-efficacy in relation to French reading. That is, they were asked to state the extent to which they believed they could successfully tackle similar texts to the ones on which they had been assessed, if presented with them again. We asked them about this for different aspects of French reading through a short questionnaire, which they completed immediately after their reading tasks. It asked them to rate how confident they were in six areas, ranging from low (*I definitely couldn't do this*) to high (*I am sure I could do this*). The six areas were as follows:

1. Read right to the end of the passage
2. Understand the main ideas in the passage
3. Understand the passage in detail
4. Work out the meaning of unknown words
5. Understand opinions expressed in the passage
6. Read the passage aloud accurately in French

For the analysis, students' responses on a four-point scale were coded as follows:

1. I definitely couldn't do this	2. I might not be able to do this	3. I could probably do this	4. I am sure I could do this
----------------------------------	-----------------------------------	-----------------------------	------------------------------

Overall self-efficacy for French reading

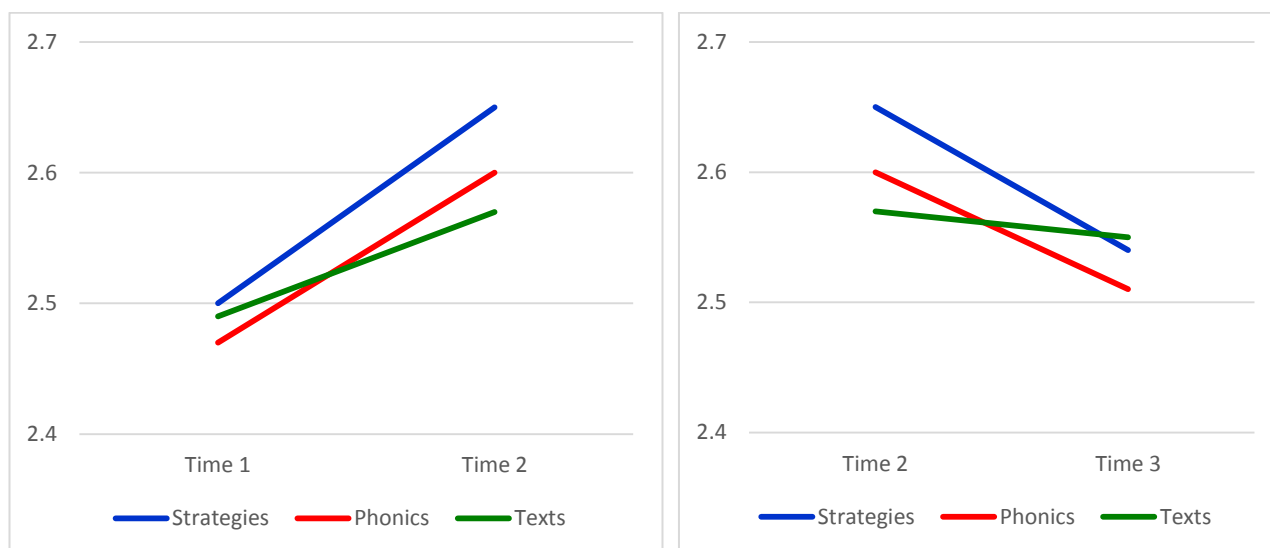
We first calculated a total self-efficacy score for each learner at each time point, by summing the scores for the six individual areas mentioned above. This score was then divided by six to give a mean total self-efficacy score out of 4. Based on this mean total self-efficacy score, on a purely descriptive level, the sample as a whole became slightly more confident in French reading between times 1 (mean 2.50) and 2 (mean 2.63). This difference was statistically significant ($p < .001$, $d = .20$) but with a small effect size. There was then slight fall in confidence at time 3, though the mean level (2.53) remained fractionally above that of time 1.

We acknowledge that these changes are very small: the time 1 to time 2 increase amounted to just over one tenth of one point on the four-point scale described above. Nevertheless, we should bear in mind that learners were asked to rate their confidence *in relation to the French texts they had just read* (and not with respect to French reading in general). The time 2 texts were designed to be more difficult than those at time 1. In other words, participants grew slightly more confident in completing their reading tasks, despite these tasks becoming considerably more challenging (and certainly more challenging, in our experience, than the reading exercises typically completed by Year 7 students).

We also looked at the changes in total self-efficacy for the three experimental groups individually. An initial analysis using a multilevel model to control for school effects (Appendix A5, Model 14) found no significant advantage for any of the groups over the others. Nonetheless, at a descriptive level (see Figure 11a), the Strategies and Phonics groups showed greater improvements than the Texts group between times 1 and 2 (albeit within the context of a very small magnitude of change overall). Furthermore, once we also controlled for prior attainment in our model, the Strategies group displayed significantly higher levels of self-efficacy than the Texts group at time 2 (Appendix A5, Model 15). Overall, therefore, there is some evidence that the Strategies group showed greater growth in self-efficacy for French reading over the course of the intervention.

By contrast, at time 3, total self-efficacy scores for both intervention groups dipped slightly below those of the Texts group (Figure 11b) – although, as has already been noted, the reduced sample size at time 3 weakens our confidence in this finding.

Figure 11: Total self-efficacy for French reading, subdivided by group, at (a) times 1-2 and (b) times 2-3¹⁰

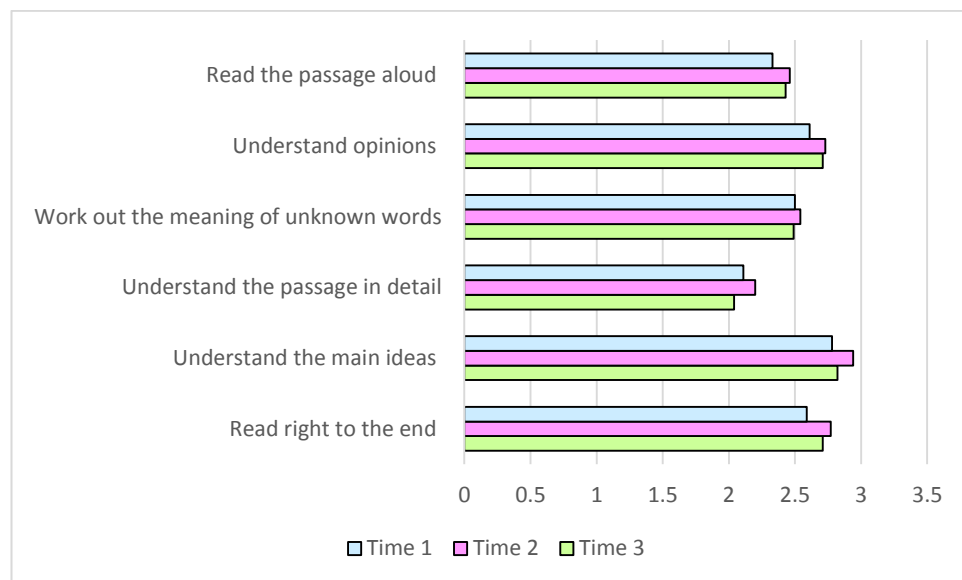


¹⁰ Here and below, we present the information in two separate graphs rather than one single graph with all three time points. This is to allow the time 1-2 graph to draw on the larger sample of students who completed the tests at these two time points (before the drop in sample size due to attrition at time 3). Note however that we did not find any significant differences between the time 2 scores according to which sample we used (the full time 1-2 sample or the reduced time 2-3 sample).

Self-efficacy in six individual aspects of French reading

We also looked at participants' scores in each individual area of self-efficacy. Looking at the sample as a whole, between times 1 and 2 there were small increases in self-efficacy for most areas (Figure 12), all statistically significant, except for *Working out the meaning of unknown words*. By time 3, however, self-efficacy levels had returned to or fallen below the time 1 levels, apart from *Understanding in detail*.

Figure 12: self-efficacy in individual areas for the whole sample



There were also some different patterns of change in self-efficacy between the three experimental groups (Table 5). For all of the changes listed in the table, the effect sizes were small.

- The strategies group showed the most widespread gains. Self-efficacy increased for all six areas between times 1 and 2 at statistically significant levels.
- For the Phonics group, self-efficacy increased for two areas between times 1 and 2 at statistically significant levels. These were: *Reading right to the end of the passage* (approaching a medium effect size) and *reading aloud*.
- For the Texts group, self-efficacy likewise increased for two areas between times 1 and 2 at statistically significant levels: *Reading right to the end of the passage* and *Understanding the main ideas*.

Pulling these findings together, we would observe that all three groups improved significantly in *Reading right to the end of the passage* (again despite the texts becoming more challenging at time 2). This change, we would hypothesize, may have arisen simply through practice in reading the longer and more challenging texts that we provided: it is something they became more used to – and, it appears, more comfortable with – as a result of all three programmes of instruction. Additionally, as we would have hoped, the Phonics group grew significantly more confident in reading aloud. The Strategies group, however, made gains across the board. They therefore seem to have experienced a more generalized increase in reading self-efficacy covering all six individual areas.

We also compared the three groups at each time point. At time 1, there were no significant differences between the groups. At time 2, however, we found small but significant differences in three areas:

- The Strategies group reported significantly higher levels of self-efficacy for *Understanding the main idea* ($p = .17$, $d = 0.20$) and *Understanding details* ($p = .015$, $d = 0.20$) compared to the Phonics group.
- The Phonics group reported significantly higher levels of self-efficacy for *Reading aloud* compared to the Texts group ($p = .015$, $d = 0.20$)

Table 5: Mean time 1 and time 2 scores for each aspect of self-efficacy, by group.

Note: Means listed in **bold type** indicate significant t1-t2 differences, as determined by Wilcoxon Signed-Rank tests; the test statistics for these significant differences are given in parenthesis.

Self-efficacy area	Time	Group Means		
		Strategies	Phonics	Texts
Read right to the end of the passage	1	2.61	2.59	2.56
	2	2.69 (p = .038, z = 2.07, d = .18)	2.85 (p < .001, z = 5.12, d = .45)	2.75 (p = .003, z = 2.99, d = .26)
Understand the main ideas in the passage	1	2.81	2.77	2.78
	2	3.02 (p < .001, z = 4.49, d = .40)	2.88	2.94 (p = .002, z = 3.15, d = .28)
Understand the passage in detail	1	2.13	2.07	2.14
	2	2.30 (p = .001, z = 3.39, d = .30)	2.16	2.15
Work out the meaning of unknown words	1	2.52	2.48	2.49
	2	2.61 (p = .026, z = 2.22, d = .20)	2.46	2.54
Understand opinions expressed in the passage	1	2.61	2.59	2.64
	2	2.81 (p < .001, z = 4.14, d = .39)	2.70	2.67
Read the passage aloud accurately in French	1	2.33	2.34	2.33
	2	2.46 (p = .015, z = 2.43, d = .22)	2.54 (p = .02, z = 2.32, d = .20)	2.35

Motivation

Learners' motivation was assessed at each time point through a questionnaire, comprising nine items which asked them to rate: (a) how much they enjoyed French; (b) how much they were looking forward to learning French next year; (c) how good they thought they were at seven aspects of learning French (reading, speaking, writing, listening, spelling, reading aloud, learning the language overall).

For the analysis, students' responses on a four-point scale were coded as follows:

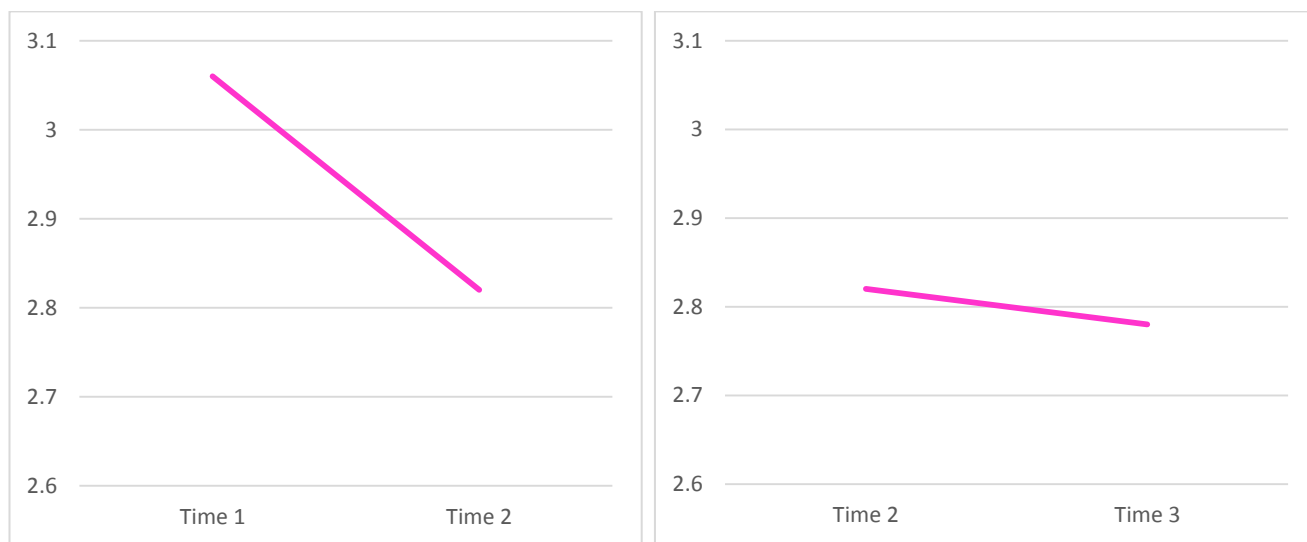
1. not at all / not good at all	2. not really / not that good	3. a bit / quite good	4. a lot / really good
---------------------------------	-------------------------------	-----------------------	------------------------

Two composite scales were computed, in order to permit the use of more powerful statistical tests than would have been possible when considering each of the nine items individually: 'Positivity' (*Enjoying French + Looking forward to French*); and 'Good at French' (the seven items relating to 'how good' they thought they were at different aspects of learning the language). These composite scales were then divided by two and seven respectively, to give mean Positivity and mean Good at French scores for each learner.

Looking at the sample as a whole, students seemed to start Year 7 with a high level of 'Positivity' for French. This then declined slightly at time 2, and declined a little further at time 3 (Figure 13). These changes were both statistically significant ($p < .001$), although very small in magnitude. (Note the large scale of the y axis in Figure 13). This pattern of a positive start to Year 7, followed by a decline in motivation across Year 7 and into Year 8, matches the findings of previous research. Nonetheless, if a score above 2.5 is taken to indicate a 'positive'

response, then learners in our study remained broadly favourably disposed towards French at time 3, and the declines they showed were only slight. In describing this overall picture, we would also caution that responses at each time point ranged from 1 to 4, indicating much variability in individual learners' overall levels of positivity.

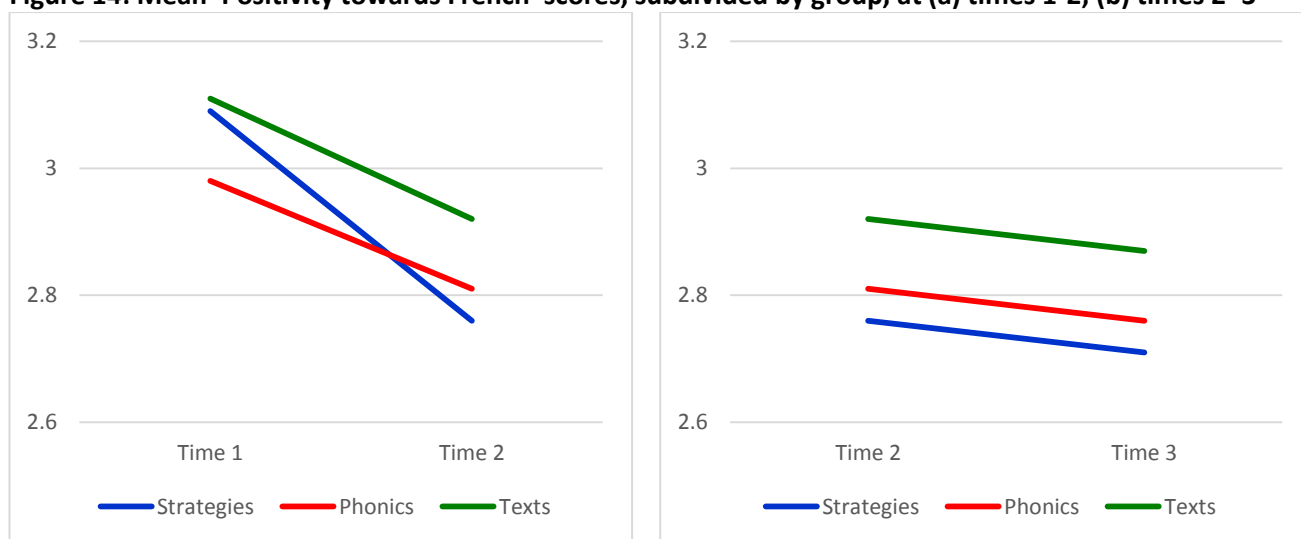
Figure 13: Mean 'Positivity towards French' scores for the sample as a whole, at (a) times 1-2; (b) times 2-3



We also looked at changes in Positivity by group. Reflecting the overall picture, all three groups started off with, and retained, a fairly high degree of positivity towards learning French, but all became less positive over time (Figure 14). Between times 1 and 2, the declines were statistically significant for all groups ($p < .001$); between times 2 and 3, only the Texts group showed a statistically significant decline ($p = .002$, though this difference is not evident in the slope of the green line in Figure 14b).

In terms of any between-groups differences, the slopes of the lines for the three groups are mostly similar, although the Strategies group appears to show a steeper decline than the others between times 1 and 2 (from 3.1 to 2.8). However, this difference between groups was not statistically significant. Therefore, we cannot say with confidence that any of our interventions was more or less successful than the others in helping students maintain a sense of positivity towards French.

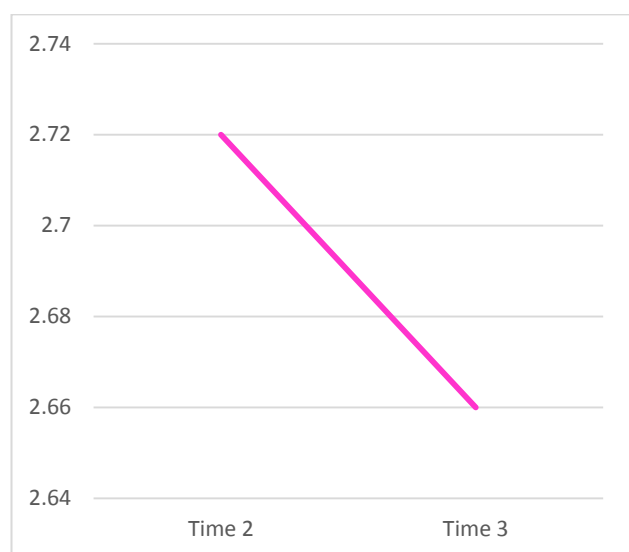
Figure 14: Mean 'Positivity towards French' scores, subdivided by group, at (a) times 1-2; (b) times 2-3



Turning to our second composite scale, on average, students' sense of being 'Good a French' amongst the sample as a whole remained constant between times 1 and 2, with an identical mean of 2.72 at both points. However, it then declined significantly at time 3, $p < .001$ (Figure 15). Again, this resonates with previous findings of a dip in

students' motivation as they move from Year 7 into Year 8. However, the magnitude of this decline was again small in our study.

Figure 15: Mean 'Good at French' scores for the sample as a whole at times 2 and 3. (Note: the mean time 1-time 2 values are not graphed, since they remained constant (at 2.72) across these two time points.)



The trends observed for the whole sample in respect of 'Good at French' were also mirrored in each of the three groups individually, with no statistically significant change between times 1 and 2 for any group, followed by a significant decline between times 2 and 3 for all groups (Strategies group: $p = .001$; Phonics, $p = .015$; Texts, $p < .001$). There were no significant differences between the groups in these changes over time.

Strategic behaviour

Learners' strategic behaviours was assessed at each time point through a questionnaire which asked them to indicate how often they used certain strategies when reading in French, choosing from *never* (1), *sometimes* (2), *often* (3) or *always* (4). The list of strategies included those taught to the Strategies group (but with different wording than had been used in the intervention materials), together with *Sounding out*. Other behaviours were also included that would be considered problematic for reading (such as *Give up*), or at least potentially problematic if not used appropriately and in conjunction with other strategies (such as *Guess from the pictures* and *Invent a meaning*).

Following Macaro and Erler (2008) we divided the list of strategies into two main groups. The first comprised what we might term positive, 'text-engagement' strategies; the second, negative or 'text-avoidance' strategies (Table 6).

At time 1, looking at the sample as a whole, students reported using all text engagement strategies at a fairly high level (mainly 'sometimes' or 'often'). At time 2, the sample as a whole showed statistically significant increases for three of these text engagement strategies, together with decreases in the following text avoidance strategies: *Wait for the teacher to explain the text*; *Ask the teacher what it means*; *Wait and see if the teacher says what it means* (all $p < .001$). Thus, there is some indication that the sample as a whole grew more independent in their French reading over the course of the intervention¹¹.

At time 3, however, there were statistically significant increases (compared with time 2) for three text avoidance strategies: *Give up*; *Ask the teacher what it means*; and *Wait and see if the teacher says what it means* (all p

¹¹ Given the high levels of reported 'text engagement' strategy use at time 1, it is also possible that there was some degree of 'ceiling effect' in operation, which may have limited the potential for any increases in the use of these strategies to be measured at time 2.

<.001). Indeed, for *Give up*, there was also a statistically significant increase compared with Time 1 ($p = .04$). In other words, there was some tendency for participants to regress from the increased independence seen at time 2. This pattern (improvement at time 2 followed by some regression at time 3) matches what was found in respect of some of the self-efficacy and motivation variables described above. Again, it is also consistent with the previous evidence of a dip in motivation and engagement as students move into Year 8.

Table 6: List of items in the strategies questionnaire, divided by group

Text engagement strategies
Try to understand each word
Try to understand new words by looking at the words around them
Go back to a word or section and double-check that it makes sense
Scan for words that look familiar and try to guess the meaning of the text from them
Think hard about words you might know
Use a process of deduction ('It can't be that, because...')
Sound out words or phrases
Use common sense to understand
When you get to a difficult sentence, read the whole of it to see if it makes sense
Look up many words in the dictionary or vocab list
Tell yourself not to give up easily
Text avoidance strategies
Wait for the teacher to explain the text
Give up
Guess from the pictures what it is all about
Invent a meaning
Ask a friend what it means
Ask the teacher what it means
Wait and see if the teacher says what it means

Looking at the strategic behaviour of the three experimental groups individually, overall the trend seemed to be for the Strategies group to show the greatest development in strategic behaviour, and for both the Strategies and Phonics groups to show greater development than the Texts group. Thus, focussing on changes between times 1 and 2, the following changes in reported strategic behaviour were observed:

1. Strategies group. There were statistically significant increases in four text-engagement strategies: *Try to understand new words by looking at the words around them* ($p = .001$); *Scan for words that look familiar and try to guess the meaning of the text from them* ($p = .001$); *Use a process of deduction – 'It can't be that, because...'* ($p = .007$); and *Use common sense to understand* ($p < .001$). There was also a significant increase in one text-avoidance strategy, *Guess from the pictures what it is all about* ($p < .001$).
2. Phonics group. There were statistically significant increases for three text-engagement strategies: *Scan for words that look familiar and try to guess the meaning of the text from them* ($p = .001$); *Try to understand new words by looking at the words around them* ($p = .001$); *Use a process of deduction – 'It can't be that, because...'* ($p = .004$). There was also a significant decrease in one text engagement strategy, *Try to understand each word* ($p < .001$).
3. Texts group. There was a significant increase for one text engagement strategy: *Scan for words that look familiar, and try to guess the meaning of the text from them* ($p = .001$). There were also significant decreases for two text engagement strategies: *Sound out words or phrases* ($p = .001$); and *Tell yourself not to give up easily* ($p = .009$).

We also compared the strategic behaviour of the three groups at each time point. At time 1, there were relatively few differences between them, although the Texts group reported more frequent use of *Sound out words or*

phrases than the Strategies group, and more frequent use of *Use a process of deduction* than the Phonics group. Thus, there was some indication that the Texts group was already more likely to use text-engaging strategic behaviour prior to the implementation of our interventions. This is unfortunate, as it represents a confounding variable.

At time 2, a larger number of significant differences was observed between the groups (summarized in Table 7). All of these significant differences relate to text-engagement strategies. Overall, there is some indication of an advantage for the Strategies group, which used a greater range of strategies (especially those relating to comprehension monitoring and checking of understanding). This advantage is clearest in relation to the Phonics group rather than the Texts group, and the Texts group also seems to show greater use of text engagement strategies compared to the Phonics group. (These findings may have been influenced by the pre-existing differences in strategic behaviour noted above in respect of the Texts group at time 1).

Table 7: Summary of between-group differences in strategic behaviour at time 2.

Strategies > Phonics	Strategies > Texts	Texts > Phonics
<ul style="list-style-type: none"> • <i>Try to understand each word</i> 	<ul style="list-style-type: none"> • <i>Guess from the pictures what it is all about</i> 	<ul style="list-style-type: none"> • <i>Try to understand each word</i>
<ul style="list-style-type: none"> • <i>When you get to a difficult sentence, read the whole of it to see if it makes sense</i> 		<ul style="list-style-type: none"> • <i>When you get to a difficult sentence, read the whole of it to see if it makes sense</i>
<ul style="list-style-type: none"> • <i>Go back to a word or section and double-check that it makes sense</i> 		<ul style="list-style-type: none"> • <i>Think hard about words you might know</i>
<ul style="list-style-type: none"> • <i>Use a process of deduction ('It can't be that, because...')</i> 		
<ul style="list-style-type: none"> • <i>Use common sense to understand</i> 		

Students' views on the programmes of instruction

We asked all three groups for their views on the pedagogical texts. Additionally, we asked the Strategies and Phonics groups for their views on the respective explicit instruction they had received.

Interestingly, whilst the pedagogical texts were designed to be challenging, around two thirds of the participating students in the sample as a whole (65.7%) felt positive about reading them. We also found statistically significant differences between the groups' mean responses. Specifically, more students in the Texts and Phonics groups said they had enjoyed reading the texts (73% and 67% of participants respectively) compared to the Strategies group (57%). Nonetheless, we would note that even in the Strategies group, well over half of the students reported that they did enjoy reading the texts.

These differences between the groups in terms of their reported enjoyment of the texts could relate to the types of tasks attached to them, rather than to the texts themselves or their content. For example, both the Strategies and Phonics groups used the texts as part of their wider programmes of explicit instruction, to practise either strategic reading or phonological decoding respectively. By contrast, the Texts group simply read the texts for meaning and answered the accompanying questions (both the simple factual ones and the more open-ended, reflective ones).

The Phonics group used these questions, too, in addition to their phonics work with the texts. As for the Strategies group, whilst they did work on the meaning of the texts, they were not asked to respond to the simple factual comprehension questions or the wider reflective questions. It is possible that some of the strategies activities, requiring a lot of self-reflection and more open-ended responses to the texts, were challenging for

some learners even if they were beneficial for self-efficacy, as has been reported in some previous studies (e.g. Souvignier & Mokhlesgerami, 2006)

Some participants also answered a follow-up question probing their reasons why they enjoyed (or did not enjoy) reading the pedagogical texts. In their responses, the word “interesting” featured in the top three most frequent words in all three groups (Strategies group: ‘interesting’ occurred 32 times out of 2188 words written in total; Phonics: 62 occurrences out of 2629 words; Texts: 42 occurrences out of 1859 words). The following comments were typical of the responses to this question:

- *[The texts] were very interesting and it was cool to be able to read in French and not just English* (Strategies group)
- *They were interesting and challenging* (Phonics group)
- *They are very interesting and overall fun to read* (Texts group)

Another open-ended question asked participants to state the main things they felt they had learnt from reading the pedagogical texts. The ‘word clouds’ in Figure 15 provide a graphical representation of the 25 most frequently-occurring words in the responses given by students in each group.

Figure 15: Word clouds showing the most frequently-mentioned ‘main things learnt from the texts’, subdivided by group

(a) Strategies group



(b) Phonics group



(c) Texts group



As can be seen, for all three groups, the key ideas “learnt”, “French” and “words” clearly predominate. Indeed, ‘words’ was the most frequent item in all cases, reflecting the fact that many students believed that they had gained vocabulary knowledge from their reading of the texts. However, there are also some interesting differences between the groups.

For the Strategies group, some words relating to strategy instruction (‘strategies’, ‘pictures’ are prominent). Students’ responses referred to the development of both strategic behaviour and cultural knowledge, as shown in the typical examples below:

- *You learn about the culture. You learn about certain French people. You learn how to use the strategies for reading. You learn new words in French.*
- *Interesting information about the person or thing the text is about and how to identify certain things and use strategies to uncover new and interesting things about the text.*

The Phonics group also contains some references specific to their programme of instruction, with ‘sounds’, ‘pronounce’, ‘pronunciation’ and ‘letters’ all featuring in the top 25 words in their responses. The following comments were typical of this group, with references not only to phonics-related learning, but also to the development of wider reading skills and cultural knowledge:

- *[You learn] how to pronounce the sounds, how to say the words, the accents.*
- *I learnt more French words I didn't know or words I knew but they were naturally put into context. I also learnt how to read French better.*
- *The sounds of separate phonics and French culture plus lots of new words.*

The Texts group mentioned neither strategies- nor phonics-related learning, focussing instead on learning about the culture. The words 'people', 'things' and 'France' were prominent. As was the case for the other two groups, they also reported that the texts had helped develop their reading skills. The following comments are characteristic and illustrate these points:

- *I have learned things about France and famous French people as well as new vocabulary.*
- *I learnt how to deduce facts and read the words in a sentence to understand what it means. I can work out the meanings of words by reading the words before and after it.*

What is particularly interesting here is that students from all three groups perceived the texts to have been a rich learning opportunity, helping them to develop their vocabulary, reading proficiency and cultural knowledge. For the Strategies and Phonics groups, the learning associated with the texts went beyond the explicit focus of their respective interventions. Further, the last quotation above from the student in the Texts group illustrates how the pedagogical texts may also have helped some students to develop their strategy use, even in the absence of explicit instruction in this area.

A sub-sample of 18 students were also interviewed individually to gain more detailed insights into their views on the pedagogical texts (and the interventions more widely). The interview data corroborated the findings of the open-ended questions described above, highlighting students' perceptions of the texts as both enjoyable and useful for developing their reading skills and knowledge of French culture. These points are illustrated in the quotations below.

As well as learning words you get to know stuff about French people and French things (Phonics group)

They helped me learn some words....a couple of months ago it (text length) would have been overwhelming (Texts group)

I enjoyed doing the texts and the sounds... the texts because I learnt more about French culture (Phonics group)

I like finding out what they do in their spare time...I'm a bit noseiy (Texts group)

because you found out about them... quite interesting... I didn't know a lot about Giroud (Texts group)

in primary school you don't really know anything... I wouldn't have dreamed before of trying to read something like that or work out what it means (Phonics group)

Finally, we sought the views of students in the Strategies and Phonics groups on the explicit teaching they had received, beyond their use of the pedagogical texts. Responses were overwhelmingly positive, with 80-85% of participants saying that the teacher presentations were useful for 'learning about the different strategies you can use' and 'learning the sounds made by letters in French' respectively.

Teachers' views on the programmes of instruction

We gathered teachers' views on both the programmes of instruction and their participation in the project as a whole, using questionnaires (completed by all teachers) and detailed interviews (a sub-sample of 14 teachers).

Two key questions in the questionnaire asked teachers whether, as a result of taking part in the project, they now 'have a better understanding of teaching reading in French' and their 'classroom teaching ... has changed'. They responded using a four point scale: 1 = strongly disagree; 2 = disagree; 3 = agree; 4 = strongly agree. Across the sample of 36 teachers as a whole, the mean response to these two questions was 3 (agree), although there were some differences by group. The mean scores for both questions is higher for the Strategies and Phonics groups, in which the teachers – as we might expect – report a greater influence on their knowledge and classroom practice. The Texts group is roughly neutral on these questions (i.e. approximately a mean score of 2.5, indicating no change as a result of the project).

In terms of teachers' views on the FLEUR teaching materials, we asked them about (a) the pedagogical texts (all teachers); (b) the phonics presentations (Phonics group only); and (c) the strategies presentations and the reflective tasks, in which students were asked to reflect on their own strategic behaviour (Strategies group only). Using the same four-point scale, ranging from strongly disagree (1) to strongly agree (4), teachers responded to a range of statements in relation to each of these materials: for example, 'Students have enjoyed using these', 'Students have learnt a lot from these' and 'I am likely to use these again with other classes in the future'.

Teachers' responses to the materials were extremely positive. The mean responses lay between 3 (agree) and 4 (strongly agree) in almost all cases, with the exception of the reflective tasks for the strategies group (mean response 2.9, so still indicating a positive view). As suggested above, it transpired that some students did not engage with these reflective tasks as well as the teachers had hoped. The explicit phonics materials received the most favourable reviews of all, with virtually all Phonics group teachers 'strongly agreeing' that they were easy to use, that students had enjoyed using them, that they had improved students' reading aloud and that they were likely to use them again with other classes in future.

The questionnaire also included an open-ended question asking teachers about the impact of the FLEUR project on their students' learning, and on their teaching more widely. A number of positive comments were made about the impact of the interventions on students, in areas such as their reading outcomes, phonological decoding, resilience, motivation, cultural knowledge and transferable skills (including for reading in English). A common theme was that teachers felt empowered to use longer, more challenging texts with Key Stage 3 learners in future. It is also interesting that there are examples in the Texts group of teachers feeling that working with the pedagogical texts had in itself improved their students' reading and strategy use, even in the absence of explicit instruction in this area. An illustrative selection from the open-ended responses is shown in below, broken down by group.

Strategies group	Phonics group	Texts group
<ul style="list-style-type: none">• Cultural aspect the kids enjoyed, which was fantastic	<ul style="list-style-type: none">• Their pronunciation has improved and their confidence in correcting others' pronunciation. Evidence that they understand GPCs rules such as final consonant.	<ul style="list-style-type: none">• I could see without teaching strategies, they were starting to work things out ... starting to develop strategies on their own
<ul style="list-style-type: none">• Senior leadership want me to cascade the strategy instruction into the department	<ul style="list-style-type: none">• Because they've accessed the longer texts they now do not complain when given a longer text, no complaints "this is hard, this is too much". Built quite a bit of resilience.	<ul style="list-style-type: none">• Think it has shown me to not shy away from challenging texts with unknown language and to use this early on in Year 7. They can do it and can learn something.

- They've got all of the core skills for those not continuing with the same language. Makes languages relevant. . I have to teach English, I find it useful in English too with Year 7. Some have low reading ages.
- Will definitely use more authentic texts in the future. They can cope with big texts.
- The cultural element of the reading was very nice as well. Very little cultural stuff in the textbooks. I liked that, it got me thinking more about the cultural element of MFL teaching, when you're on autopilot it's easy to overlook.
- I'm very interested in phonics, for me it's fundamental. The resources I have been supplied with are a treasure for me that I will keep to use for years to come.
- Huge difference in that class's attainment for reading at the end of year.....Tackling texts that were more difficult for them, they have really improved even from spring and summer assessments. Weaker students have made huge improvement.

There were also a smaller number of more critical comments. For example, one teacher noted that the texts were difficult for lower attaining students (though this contrasts with the third quotation in the Texts group above). One other noted that it became harder to maintain some students' motivation in the strategy instruction as time went on: "a third lost interest and some remained really keen".

Discussion and conclusions

In this section, we return to our research questions and offer a brief summary of the evidence in relation to these. As a reminder, our research questions were:

1. To what extent does a programme of explicit instruction in (a) reading strategies or (b) phonics lead to improvements in French reading proficiency amongst Year 7 MFL students, as compared to a programme of reading challenging texts?
2. To what extent do any of the three programmes of instruction lead to improvements in: (a) students' French phonological decoding proficiency; (b) their strategic behaviour when reading in French; (c) their vocabulary knowledge; and (d) their motivation for learning the language?
3. Is one of the programmes of instruction more effective than the others on any of the above measures?
4. What are teachers' and students' views on the three programmes of instruction?

Given the attrition that we experienced at time 3, we consider our principal evidence in relation to these questions from the data at times 1 and 2. Further, the time 3 data collection took place after the long Summer holiday, followed by a whole school term of teaching under different instructional approaches, and in many cases with a different teacher. Because of this, it is unsurprising that some of the positive changes that we saw between times 1 and 2 were not maintained (or even went into reverse) at time 3. This implies that for any positive effects of a programme of instruction to be sustained in the longer term, then that instruction must also continue for a longer period of time.

Impact on reading comprehension

Turning first to the impact of the programmes of instruction on reading comprehension, once we had taken account of school effects in our analysis, we did not find any evidence that any of the programmes of instruction was more effective in improving reading outcomes than the others, either in the immediate term or in the longer term (6 months after the end of the interventions). Neither explicit Strategies instruction, nor explicit Phonics instruction (as implemented in our study) provided a 'magic solution' for developing students' foreign language reading, over and above the provision of appropriately challenging, engaging texts.

There are a number of possible reasons for the lack of significant differences between the groups in our study. It is instructive here to compare our findings to those of Macaro and Erler (2008), where the combined strategies-and-phonics intervention did lead to significantly better outcomes than the control group.

The first point to note relates to the nature of the interventions. One of the limitations we identified in Macaro and Erler's (2008) study was the fact that their control group, in following its usual curriculum, had no access to the more challenging texts which were used as part of the strategies intervention. It is therefore possible (although the authors argue against this) that their Strategies group did better simply through engaging with more challenging and interesting texts than those routinely included in the Key Stage 3 curriculum. To address this limitation, all three groups in our study worked with the same 'pedagogical texts'. This may in itself have helped students get better at reading in French. This interpretation is supported by both students' and teachers' responses in the post-intervention questionnaires and interviews: these highlighted the rich learning opportunities afforded by the texts.

In this regard, we would also note that all three groups showed a significant improvement in reading outcomes, when we directly compared their scores on common items across times 1 and 2. Of course, some degree of progress could reasonably be expected over this time period in any case – although, given other evidence concerning MFL students' low motivation and limited progress in Key Stage 3, this cannot be taken for granted. Unfortunately, because Macaro and Erler (2008) used different reading tests at times 1 and 2, we do not know how much progress, if any, their control group made in reading over the course of their intervention. Neither was it possible, in our own study, to include a 'true control' group, having no exposure to the pedagogical texts, which would have allowed us to make stronger causal claims about the effects of these texts on students' outcomes¹².

A second possible reason for the lack of significant between-group differences in our study is that a combined strategies-and-phonics intervention (as implemented in Macaro and Erler, 2008) may be necessary in order to have a larger impact on reading outcomes. Interestingly, this view was put forward to us informally by one of our participating teachers, before hearing about our findings. 'Sounding out' (which formed part of Macaro and Erler's intervention) may be a key strategy to include in any explicit strategy instruction programme. Conversely, this strategy presupposes a basis of French GPC knowledge in order for it to operate effectively, thereby implying the need for some phonics instruction alongside the strategy instruction.

Third, it may also be that a longer intervention is needed in order for a greater impact to be obtained. The intervention in Macaro and Erler's (2008) study lasted 15 months, using a 'drip feed' approach of around 10 minutes' instruction per week. Our own intervention, by contrast, was shorter (16 weeks) and more intensive (20-30 minutes per week).

Finally, we would note that our study used multi-level statistical tests which take account of the fact that our participants were 'clustered' within schools. When we conducted more traditional statistical tests without this adjustment, we did find significant differences in reading outcomes between the groups (with an advantage for the Phonics group). It would be interesting to re-analyse the data from Macaro and Erler (2008), where participants were clustered in six different schools, using a multilevel model, to see whether significant differences between the intervention and control groups remain.

We always intended our study to complement Macaro and Erler's (2008) study. It was designed to isolate the effects of explicit phonics instruction and explicit strategies instruction, which we could then compare with their integrated strategies-and-phonics programme. On this basis, we would still conclude that overall, an integrated programme is likely to be the most effective approach for developing students' reading comprehension in Modern Foreign Languages classrooms. Nonetheless, it remains possible that getting students to engage with

¹² Our sample was widely dispersed geographically (much more so than we had originally anticipated, due to challenges in recruiting schools to participate) and we considered that adding a further group of schools would have made data collection unmanageable, given the size of the research team – particularly given the need to administer tests 'blind to condition' and within a short time window at each data collection point.

appropriately challenging and engaging texts may in itself be beneficial for their reading outcomes. The evidence suggests to us that texts such as the ones used in our study may form a valuable contribution to any integrated programme of MFL reading instruction. Further research is needed to test this hypothesis, for example by comparing a Texts group, as implemented in this study, to a true control group (and perhaps also to an 'authentic texts' group, since our 'pedagogical texts' were also designed to provide rich opportunities for learning and practising particular strategies and grapheme-phoneme correspondences in the foreign language).

Other linguistic outcomes: phonological decoding and vocabulary knowledge

We also looked at the impact of our three programmes of instruction on students' phonological decoding and vocabulary knowledge.

On the decoding tests, as expected, we observed an advantage for the Phonics group in terms of the progress made between times 1 and 2. The Phonics group, on average, scored 6-7 percentage points higher than both the Strategies and the Texts groups at time 2, and this advantage was on the cusp of statistical significance. The fact that a clearer statistical difference was not found may be because our pen-and-paper decoding test (the SALT) was not sensitive enough to fully capture the progress that students had made; a Reading Aloud Task would have been better, but was unfortunately not feasible with such a large sample. When we additionally took account of students' prior academic attainment in our statistical model, the advantage for the Phonics group was statistically significant, even using our less sensitive SALT. Further, whilst all groups in our study (including those not receiving explicit phonics instruction) did make significant progress in decoding between times 1 and 2, this progress was of a smaller magnitude for the Strategies and Texts group compared to the Phonics group. These findings echo previous, smaller-scale research in the MFL context (e.g. Woore, 2011). Looking at the available evidence as a whole, we therefore consider it likely that explicit Phonics instruction is beneficial for developing students' phonological decoding (and perhaps even necessary, in order for any meaningful progress to be made).

In respect of vocabulary knowledge, we found significant differences between the groups, even in multilevel models which controlled for school effects. Specifically, in our basic model, there was a significant advantage for the Phonics group over the Texts group; and in a model which additionally controlled for students' prior attainment, both the Strategies and the Phonics group showed a significant advantage over the Texts group. Further, whilst all three groups showed significant increases in estimated vocabulary knowledge, the increase for the Strategies group was almost double that for the Texts group, and the increase for the Phonics group was greater still (over 2.5 times higher than that of the Texts group).

We would hypothesize that both the Strategies and the Phonics groups engaged with the Pedagogical Texts on a deeper level. These texts, which are longer than those that Year 7 learners usually engage with, inevitably contain a substrate of high frequency vocabulary (alongside whatever specialist vocabulary is associated with their particular topic). Engaging with these texts may therefore have led to some gains in high frequency vocabulary knowledge, as measured on the X_Lex test.

Further, the findings of a particular advantage for the Phonics group echo emerging evidence from other contexts, which indicates that decoding proficiency may be important for vocabulary learning in a foreign language (e.g. Hamada & Koda, 2008, 2011). Indeed, there is some very recent experimental research (with a sample of learners of English in China) that explicit phonics instruction can lead to improvements in vocabulary learning (Li, in preparation). Given that vocabulary knowledge is absolutely central to language learning and is known to correlate with proficiency in all the four skills (Milton, 2013), this is potentially an important finding. It highlights the fact that, in a foreign language context, the rationale for phonics instruction may be different from that in primary school classrooms, where children are learning to read in their native language and sounding words out is a key tool for discovering their meanings.

Strategic behaviour

The picture in relation to participants' strategy use is somewhat complex, requiring us to undertake further, more qualitative and fine-grained analyses (including in relation to students' prior academic attainment). However, the analyses in this report provide some indication of an advantage for the Strategies group at time 2, particularly in comparison to the Phonics group. In terms of changes observed in each group's strategy use over time, the Strategies group again appears to have shown the most positive developments between times 1 and 2 (in terms of greater text engagement), followed by the Phonics group – which did not receive any explicit strategy instruction, yet did engage in detail with the pedagogical texts. The Texts group, by contrast, appears to have shown little increase in text engagement (and indeed there was arguably a slight decline in this group). On the other hand, there was also evidence that the Texts group (a) used several text-engaging strategies significantly more than the Phonics group at time 2; but also (b) may already have shown a greater likelihood to use such strategies than the other two groups at time 1 (before the start of our interventions).

Overall, we would conclude that explicit strategy instruction is probably beneficial for the development of positive, text-engaged strategic behaviour; but that it is also possible that engaging with more challenging texts can, in itself, promote the development of strategic behaviour, even in the absence of explicit strategy instruction (cf. Yuill & Oakhill, 1988). This latter view is also supported by some of the responses to the teacher and student questionnaires. For example, we cited one student in the Texts group who reported that, as a result of reading the pedagogical texts, "I can work out the meanings of words by reading the words before and after it"; and a teacher in the Texts group who said that "I could see without teaching strategies, they were starting to work things out ... starting to develop strategies on their own".

Attitudinal outcomes: self-efficacy and motivation

In terms of self-efficacy, the most widespread gains appeared to be for the Strategies group, who showed significant improvements in all six of the individual areas that we asked about. Furthermore, once we controlled for students' prior academic attainment, the Strategies group displayed significantly higher levels of overall self-efficacy than the Texts group at time 2. This suggests an important potential benefit of the strategies instruction, even though it was not matched by a significant advantage in terms of reading comprehension scores (at least within the time period of our study). Students who feel more positive and more confident about reading are likely to read more, and thus show greater improvements in the longer term.

We also found that all three groups grew significantly more confident that they could *Read right to the end of a passage* in French, despite the associated texts growing more challenging over time. Once again, it may be that this change derives from the greater practice that participants in all groups had in reading the Pedagogical French texts which formed part of the interventions. Finally, as might have been expected, the Phonics group reported significantly greater confidence in reading aloud as a result of the intervention.

We did not find evidence of any significant advantage for any of the programmes of instruction in terms of their impact on students' motivation. Unfortunately, whilst all three groups showed high levels of 'positivity towards French' at time 1, they showed a significant decline by time 2, although the magnitude of the decline was very small. Their sense of being 'good at French', meanwhile, remained at its original level (a little above the neutral position) between times 1 and 2, but then declined significantly at time 3. As we noted earlier, this resonates with previous findings of a dip in students' motivation as they move from Year 7 into Year 8 (Coleman et al, 2007), although in our case the decline was very small.

Other views on the interventions

The discussion above has highlighted that the pedagogical texts themselves, which participants in all three groups worked with as part of their programmes of instruction, may have had a beneficial effect on a range of learners' outcomes. As we saw, this was echoed in various comments from participating students and teachers in their

post-intervention questionnaires and interviews. Students and teachers also commented positively on the cultural aspects of the texts, saying how much they had enjoyed this and learnt from it.

We understand the arguments for using simple reading texts in a foreign language context – to avoid overwhelming students, to allow them to practise familiar language, and to ensure that they recognize sufficient vocabulary for their existing reading skills to come into play (Schmitt et al., 2011; Grabe & Stoller, 2011). There is sound research evidence underpinning such a view. However, our concern is that if the *only* reading texts that students encounter in a foreign language are composed of simple, familiar language, then this will eventually take its toll on their levels of engagement. What is the point of reading in a foreign language (so much more effortful than reading in a native language!) if the material you read is trivial and unrewarding? Therefore, as we have argued previously (Macaro et al., 2015; Woore, 2014), we believe that there is merit in providing students with longer, more challenging and interesting texts at least some of the time. In the words of the UK Literacy Association – referring to native language reading instruction, but we believe equally applicable to a foreign language context, “If we want England’s children to get better at reading and to do more of it, we have to give them a diet that is attractive, nutritious and satisfying” (UKLA, 2010).

This approach echoes the ambitious targets for foreign language reading set out in the new National Curriculum (DfE, 2013), which invites teachers to engage students with a range of texts, including literary ones and “great literature”, in order “to stimulate ideas, develop creative expression and expand understanding of the language and culture”. We believe that reading in MFL classrooms has too often been unambitious and unexciting. We further believe that our study has demonstrated that the use of challenging (and thereby interesting) texts is not only possible with young, near-beginner learners in an MFL context, but also potentially of great benefit to them. In the words of the teacher quoted above, “it has shown me to not shy away from challenging texts with unknown language and to use this early on in Year 7. They can do it and can learn something”.

Key recommendations

Drawing on the findings of the current study, and combining these with previous evidence in this area, we offer the following key recommendations for MFL teachers in relation to the teaching of reading at Key Stage 3:

- **Expectations regarding the kinds of written texts that beginner learners of French can tackle should be raised.** Our study shows that it is possible to use challenging, engaging texts, covering cultural topics of genuine interest to students, even with beginner learners. Further, our study suggests that the use of such texts can promote several aspects of reading and language development, as well as being popular with both teachers and learners. Therefore, we believe that including such texts in the Year 7 curriculum will be beneficial for students’ linguistic and motivational development. They should not be restricted to a sole diet of the shorter, simpler and more predictable kinds of texts that are traditionally associated with this age group.
- **An integrated approach to French reading instruction – combining explicit instruction in both Strategies and Phonics with the use of appropriately challenging, engaging texts – is more likely to be beneficial than any of these approaches in isolation.** All three groups in our study made significant progress in reading French. However, there was evidence of an advantage for the Phonics group in phonological decoding, and for both the Strategies and the Phonics groups (most clearly for the latter) in vocabulary learning. There was also some evidence that the Strategies group made more progress in strategic behaviour and self-efficacy. Bringing these benefits together, alongside the use of challenging texts, is likely to lead to more engaged learners who have the linguistic knowledge, strategic proficiency, and resilience to comprehend written texts and hence benefit from them.
- **These approaches need to be used in the long term, beyond Year 7, to retain their benefits, and they should be integrated into a focus on overall literacy development.**
- We consider phonological decoding (the ability to read words aloud accurately and fluently) to be a foundation literacy skill in a foreign language. It cannot be assumed that students will develop this skill spontaneously in French. **It is highly likely that explicit phonics instruction is beneficial, and indeed may be necessary, for many MFL students to learn to decode in French.**

- **Further development work should be undertaken with teachers to enable them to adapt their existing Schemes of Work**, to incorporate an integrated approach to reading instruction for learners with a range of different prior attainment levels, across a range of year groups and in a range of languages.

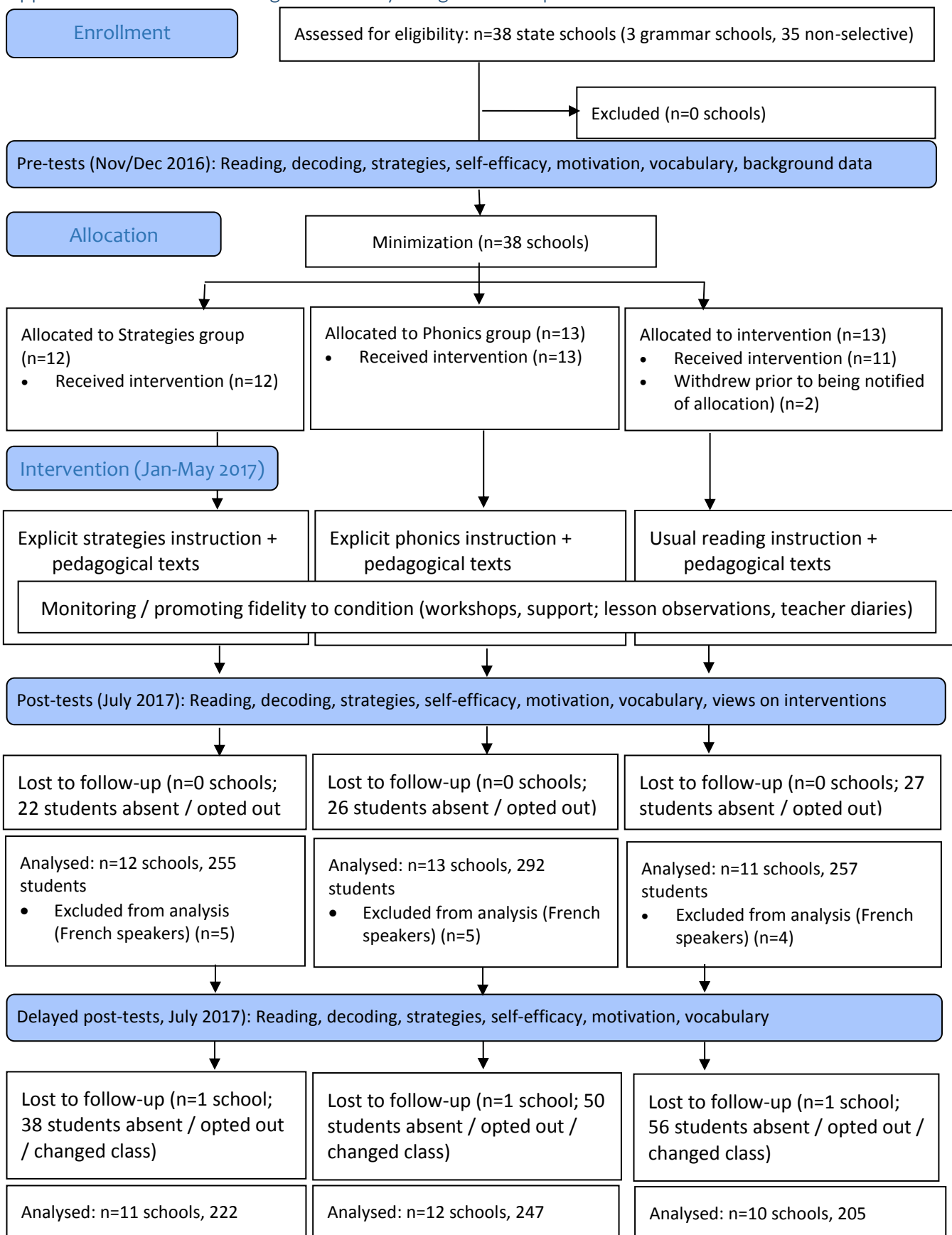
References

- Andon, N. and Wingate, U. (2012) *The Contribution of Textbooks to Sustaining Motivation for Second Language Learning*. Paper presented at 8th Annual Conference of the British Association for Applied Linguists Language Learning and Teaching SIG, Oxford University, 4th-5th July 2012.
- Ardasheva, Y., Wang, Z., Adesope, O.O. and Valentine, J.C. (2017) 'Exploring effectiveness and moderators of language learning strategy instruction on second language and self-regulated learning outcomes'. *Review of Educational Research*, 87(3), pp. 544-582.
- Bialystok, E. (2007) 'Cognitive effects of bilingualism: how linguistic experience leads to cognitive change'. *International Journal of Bilingual Education and Bilingualism*, 10(3), pp. 210-223.
- Board, K. and Tinsley, T. (2015) *Language Trends 2014/15: The state of language learning in primary and secondary schools in England*.
- CBI (Confederation of British Industry) (2012) *Learning to grow: what employers need from education and skills. Education and skills survey 2012*.
- Coleman, J. A., Galaczi, A. and Astruc, L. (2007) 'Motivation of UK school pupils towards foreign languages: a large-scale survey at key Stage 3'. *Language Learning Journal*, 35(2), pp. 245-281.
- DfE (Department for Education) (2013) *National Curriculum in England: Languages programmes of study*.
- Erlar, L. (2003) *Reading in a Foreign Language – Near-Beginner Adolescents' Experiences of French in English Secondary Schools*. Unpublished DPhil Thesis, Oxford University.
- Erlar, L. (2004) 'Near-Beginner Learners of French Are Reading at a Disability Level'. *Francophonie*, 30, pp. 9-15.
- Erlar, L. and Macaro, E. (2012) 'Decoding Ability in French as a Foreign Language and Language Learning Motivation'. *The Modern Language Journal*, 95(4), pp. 496-518.
- European Commission (2012) *First European Survey on Language Competences: Executive Summary*.
- Grabe, W. and Stoller, F. L. (2011) *Teaching and researching reading*. London: Longman.
- Graham, S., Courtney, L., Tonkyn, A. and Marinis, T. (2016) 'Motivational trajectories for early language learning across the primary–secondary school transition'. *British Educational Research Journal*, 42(4), pp.682-702.
- Hamada, M. and Koda, K., (2008) 'Influence of first language orthographic experience on second language decoding and word learning'. *Language Learning*, 58(1), pp. 1-31.
- Hamada, M. and Koda, K. (2011) 'The role of the phonological loop in English word learning: A comparison of Chinese ESL learners and native speakers'. *Journal of Psycholinguistic Research*, 40(2), pp.75-92.
- Macaro, E. and Erlar, L. (2008) 'Raising the achievement of young-beginner readers of French through strategy instruction'. *Applied Linguistics*, 29(1), pp. 90-119.
- Macaro, E., Graham, S. and Woore, R. (2015) *Improving Foreign Language Teaching: Towards a research-based curriculum and pedagogy*. Abingdon: Routledge.
- McGowan, P. and Turner, M. (1994) 'Raising reading attainment in modern languages', in A. Swarbrick (ed.), *Teaching modern foreign languages in secondary schools: a reader*. London: Routledge. Pp. 125-137.

- Meara, P.M. and Milton, J., 2003. *X_Lex. The Swansea vocabulary levels test*. Newbury: Express.
- Moore, L., Graham, A. and Diamond, I. (2003) 'On the feasibility of conducting randomised trials in education: case study of a sex education intervention'. *British Educational Research Journal*, 29(5), pp. 673-689.
- Ofsted (Office for Standards in Education, Children's Services and Skills) (2011) *Modern Languages: Achievement and challenge 2007–2010*. Manchester: Ofsted.
- Plonsky, L. (2011) 'The Effectiveness of Second Language Strategy Instruction: A Meta-analysis'. *Language learning*, 61(4), pp. 993-1038.
- Sammons, P., Toth, K. and Sylva, K. (2015) *Subject to background: what promotes better achievement for bright but disadvantaged students?* Available from www.suttontrust.com.
- Souvignier, E. and Mokhlesgerami, J. (2006) 'Using self-regulation as a framework for implementing strategy instruction to foster reading comprehension'. *Learning and instruction*, 16(1), pp. 57-71.
- Spinath, B. and Steinmayr, R. (2008) 'Longitudinal analysis of intrinsic motivation and competence beliefs: Is there a relation over time?' *Child Development*, 79(5), pp. 1555-1569.
- Swan, M. (2008) 'Talking Sense about Learning Strategies'. *RELC Journal*, 39(2), pp. 262-273.
- Tinsley, T. and Doležal, N. (2018) *Language Trends 2018: Language teaching in primary and secondary schools in England. Survey Report*. British Council.
- Torgerson, C. J. and Torgerson, D. J. with Styles, B. (2013) *Randomised trials in education: An introductory handbook* (draft in progress).
- Woore, R. (2011) *Investigating and developing beginner learners' decoding proficiency in second language French: an evaluation of two programmes of instruction*. Unpublished PhD thesis, University of Oxford.
- Woore, R. (2014) 'Developing reading and decoding in the MFL classroom', in P. Driscoll, E. Macaro and A. Swarbrick (eds.) *Debates in Modern Languages Education*. London: Routledge, pp. 81-95.

Appendix

Appendix 1: Consort flow diagram of study design and sample



Appendix 2: Details of statistical tests

Appendix 2.1: Multilevel models where DV = time 2 reading scores on items common across times 1 and 2.

To examine the effects of the interventions on reading outcomes, we conducted a two-level multilevel linear regression model, with individual participants at level 1 and school at level 2. Participants' t2 scores (on the common test items) were the dependent variable. Their t1 scores on those same items were entered as a covariate (allowing us to control for participants' baseline scores when examining their performance after the interventions). We also used two dummy variables (one for the Strategies group and one for the Phonics group), to compare these groups' performance with that of the Texts group. This allowed us to see whether the explicit Phonics and Strategies instruction had made any difference to reading outcomes over and above the use of the pedagogical texts alone. The model parameters are shown in Table A1, Model 2. (Model 1 in this table shows the parameters for a fixed-effects only model in which no account was taken of school-level effects).

The time 2 scores showed significant variance in intercepts between schools, $\text{var}(u_{0j})=11.13$, $\chi^2(1) = 167.191$, $p < .001$. In other words, a significant portion of the variance (26.8%) in time 2 scores, after controlling for time 1 scores, was accounted for at the school level. Unsurprisingly, time 1 scores strongly and significantly predicted time 2 scores, $F(1, 796.27) = 761.677$, $p < .001$. However, being in the strategies group did not have a significant effect on outcomes (t2 scores) compared to the Texts group, $F(1, 35.890) = .328$, $p = .570$ and nor did being in phonics group, $F(1, 35.454) = 1.333$, $p = .356$.

In a further model (Model 3, Table A1), we also added prior attainment (in the form of z scores based on participants' average CAT scores or average Key Stage 2 SAT scores). This significantly improved the overall model fit. Of the variance in time 2 reading comprehension scores, 28.3% was accounted for at the school level. There was a significant effect of both time 1 score, $F(1, 617.717) = 545.92$, $p < .001$, and of prior attainment, $F(1, 622.637) = 28.90$, $p < .001$, on time 2 reading outcomes. However, there remained no significant effect on outcomes of being in either the Strategies group, $F(1, 28.238) = 0.156$, $p = .696$, or the Phonics group, $F(1, 27.809) = 0.426$, $p = .519$, compared to being in the Texts group.

Note that prior attainment data was not available for all participants, and so the sample size for model 6 (n=695) is lower than that for models 4 and 5 (n=804).

Table A1: Models of the effects of the interventions on time 2 reading outcomes (for common items across times 1 and 2)

fixed effects	Model 1			Model 2			Model 3		
	b	SE	95% CI	b	SE	95% CI	b	SE	95% CI
t1 reading score	0.84***	0.07	0.78, 0.89	0.76***	0.03	0.71, 0.82	0.69	0.03	0.64, 0.75
Strategies group	0.65	0.57	-0.47, 1.77	0.85	1.48	-2.16, 3.86	1.29	3.27	-5.41, 7.99
Phonics group	1.86	0.55	0.78, 2.94	1.67	1.45	-1.27, 4.62	2.03	3.12	-4.35, 8.42
Prior attainment (z score)							2.83	0.53	1.80, 3.86
random effects				σ^2	SE	95% CI	σ^2	SE	95% CI
school				11.13***	2.98	6.59, 18.80	40.12***	12.05	22.27, 72.27
model fit									
-2LL	5272.0			5104.8			4741.4		
$\Delta \chi^2 (p)$				167.2***			363.4***		

* $p < .05$; ** $p < .01$; *** $p < .001$

Appendix 2.2: Multilevel models where DV = overall time 2 reading scores on all items.

In a two-level linear model with individual participants at level 1 and schools at level 2, the overall time 2 scores showed significant variance in intercepts between schools, $\text{var}(u_{0j})=45.24$, $\chi^2(1) = 199.759$, $p < .001$. In other words, a significant portion of the variance (29.6%) in time 2 scores, after controlling for time 1 scores, was accounted for at the school level. Time 1 scores significantly predicted time 2 scores, $F(1, 44.82) = 10.57$, $p = .002$. However, being in the Strategies group did not have a significant effect on t2 scores compared to the Texts group, $F(1, 36.24) = .003$, $p = .958$ and nor did being in the Phonics group, $F(1, 35.77) = 0.094$, $p = .761$. The model parameters are shown in Table A2, Model 5. (In the same table, Model 4 shows the parameters for a fixed effects only model, without allowing intercepts to vary across schools).

A further model (Model 6, Table A2) added prior attainment as a fixed main effect. This resulted in significantly improved model fit. Of the overall variance in time 2 scores, 28.3% was accounted for at the school level. There was a significant effect on overall time 2 reading scores of both time 1 reading scores, $F(1, 617.72) = 545.92$, $p < .001$, and of prior attainment, $F(1, 622.64) = 28.90$, $p < .001$. However, there was still no significant effect of being in either the Strategies group, $F(1, 28.24) = 0.156$, $p = .696$, or the Phonics group, $F(1, 27.81) = 0.426$, $p = .519$, compared to being in the Texts group.

Note that prior attainment data was not available for all participants, and so the sample size for model 6 ($n=693$) is lower than that for models 4 and 5 ($n=804$).

Table A2: Models of the effects of the interventions on time 2 reading outcomes (overall scores)

fixed effects	Model 4			Model 5			Model 6		
	b	SE	95% CI	b	SE	95% CI	b	SE	95% CI
t1 reading score	0.83**	0.02	0.79, 0.87	0.76	0.02	0.71, 0.81	0.69	0.03	0.64, 0.75
Strategies group	-0.79	1.10	-2.95, 1.37	-0.16	2.97	-6.18, 5.86	1.29	3.27	-5.41, 7.99
Phonics group	0.92	1.06	-1.16, 2.99	0.89	2.90	-5.00, 6.78	2.03	3.12	-4.35, 8.42
Prior attainment (z score)							2.83	0.53	1.80, 3.86
random effects				σ^2	SE	95% CI	σ^2	SE	95% CI
school				45.24	11.87	27.06, 75.65	40.12	12.05	22.27, 72.27
model fit									
-2LL	6370.5			6170.8			4741.4		
$\Delta \chi^2(p)$				199.8***			1429.4***		

* $p < .05$; ** $p < .01$; *** $p < .001$

Appendix 2.3: Multilevel models where DV = time 2 decoding scores

In a two-level linear model with individual participants at level 1 and schools at level 2, participants' overall time 2 scores on the decoding test (SALT) showed significant variance in intercepts between schools, $\text{var}(u_{0j})=32.64$, $\chi^2(1) = 101.28$, $p < .001$. In other words, a significant portion of the variance (19.2%) in time 2 scores, after controlling for time 1 scores, was accounted for at the school level. Time 1 scores significantly predicted time 2 scores, $F(1, 794.837) = 142.939$, $p < .001$. Being in the Strategies group did not have a significant effect on t2 decoding scores compared to the Texts group, $F(1, 36.04) = .072$, $p = .791$. Nor did being in the Phonics group, although the p value closely approached significance, $F(1, 35.74) = 4.02$, $p = .053$. The model parameters are shown in Table A3, Model 8. (In the same table, Model 7 shows the parameters for a fixed effects only model, without allowing intercepts to vary across schools).

A further model (Model 9, Table A3) was conducted with the addition of prior attainment as a fixed main effect. This resulted in significantly improved model fit. Of the overall variance in time 2 scores, 20.5% was accounted for at the school level. There was a significant effect on overall time 2 decoding scores of both time 1 decoding scores, $F(1, 605.39) = 89.47$, $p < .001$, and of prior attainment, $F(1, 609.68) = 53.57$, $p < .001$. There was still no significant effect of being in the Strategies group compared to the Texts group, $F(1, 28.07) = 0.009$, $p = .926$. However, there was a significant effect of being in the Phonics group compared to the Texts group, $F(1, 27.76) = 4.74$, $p = .038$.

Note that prior attainment data was not available for all participants, and so the sample size for model 9 ($n=693$) is lower than that for models 7 and 8 ($n=800$).

Table A3: Models of the effects of the interventions on time 2 reading outcomes (overall scores)

	Model 7			Model 8			Model 9		
fixed effects	b	SE	95% CI	b	SE	95% CI	b	SE	95% CI
t1 decoding score	0.37***	0.03	0.32, 0.42	0.29***	0.02	0.24, 0.34	0.25***	0.03	0.20, 0.30
Strategies group	-1.15	1.16	-3.41, 1.12	-0.70	2.62	-6.00, 4.60	0.28	2.99	-5.84, 6.40
Phonics group	5.15	1.12	2.94, 7.36	5.13	2.56	-0.06, 10.33	6.20	2.85	0.36, 12.03
Prior attainment (z score)							4.44***	0.61	3.25, 5.63
random effects				σ^2	SE	95% CI	σ^2	SE	95% CI
school				32.65	9.23	18.75, 56.82	31.55	10.07	16.88, 58.99
model fit									
-2LL	6377.5			6276.3			4755.0		
$\Delta \chi^2(p)$				101.3***			1521.3***		

* $p < .05$; ** $p < .01$; *** $p < .001$

Appendix 2.4: Multilevel models where DV = time 2 vocabulary scores

In a two-level linear model with individual participants at level 1 and schools at level 2, participants' overall time 2 scores on the vocabulary test showed significant variance in intercepts between schools, $\text{var}(u_{0j})=3108.81$, $\chi^2(1) = 9.9$, $p < .01$. In other words, a small but significant portion of the variance (5.2%) in time 2 vocabulary scores, after controlling for time 1 scores, was accounted for at the school level. Time 1 scores significantly predicted time 2 scores, $F(1, 646.96) = 41.39$, $p < .001$. Being in the Strategies group did not have a significant effect on t2 vocabulary scores compared to the Texts group, $F(1, 35.83) = 2.91$, $p = .097$. However, being in the Phonics group did have a significant effect on t2 vocabulary scores, $F(1, 34.35) = 7.16$, $p = .011$. The model parameters are shown in Table A4, Model 11. (In the same table, Model 10 shows the parameters for a fixed effects only model, without allowing intercepts to vary across schools).

A further model (Model 12, Table A4) was conducted with the addition of prior attainment as a fixed main effect. This improved the model fit significantly. 3.3% of the overall variance in time 2 scores was accounted for at the school level. There was a significant effect on overall time 2 vocabulary scores of both time 1 vocabulary scores, $F(1, 500.93) = 30.30$, $p < .001$, and of prior attainment, $F(1, 263.49) = 4.18$, $p = .042$. There was now a significant effect both of being in the Strategies group compared to the Texts group, $F(1, 28.27) = 7.39$, $p = .011$, and of being in the Phonics group, $F(1, 26.31) = 5.57$, $p = .026$.

Note that prior attainment data was not available for all participants, and so the sample size for model 12 ($n=501$) is lower than that for models 10 and 11 ($n=647$).

Table A4: Models of the effects of the interventions on time 2 reading outcomes (overall scores)

	Model 10			Model 11			Model 12		
fixed effects	b	SE	95% CI	b	SE	95% CI	b	SE	95% CI
t1 vocab score	0.28	0.04	0.20, 0.37	0.28	0.04	0.19, 0.36	0.28	0.05	0.18, 0.38
Strategies group	54.41	24.28	6.74, 102.08	57.10	33.47	-10.80, 124.99	97.60	35.90	24.11, 171.10
Phonics group	80.40	23.33	34.59, 126.21	86.92	32.49	20.92, 152.92	79.32	33.61	10.28, 148.36
Prior attainment (z score)							26.73	13.07	0.99, 52.46
random effects				σ^2	SE	95% CI	σ^2	SE	95% CI
school				3108.81	1499.64	1207.78, 8002.04	1950.37	1425.99	465.34, 8174.57
model fit									
-2LL	8957.1			8947.1			6930.3		
$\Delta \chi^2(p)$				9.9**			2016.8***		

* $p < .05$; ** $p < .01$; *** $p < .001$

Appendix 2.5: Multilevel models where DV = time 2 total self-efficacy scores

The time 2 total self-efficacy scores showed significant variance in intercepts between schools, $\text{var}(u_{0j}) = 0.49$, $\chi^2(1) = 14.664$, $p < .001$. In other words, a small but significant portion of the variance (5.8%) in total time 2 self-efficacy scores, after controlling for the time 1 scores, was accounted for at the school level. Time 1 scores strongly and significantly predicted time 2 scores, $F(1, 711.99) = 283.05$, $p < .001$. However, being in the strategies group did not have a significant effect on outcomes (t2 scores) compared to the Texts group, $F(1, 35.23) = 3.14$, $p = .085$, and nor did being in the Phonics group, $F(1, 35.15) = 0.05$, $p = .825$. The model parameters are shown in Table A5, Model 14. (In the same table, Model 13 shows the parameters for a fixed effects only model, without allowing intercepts to vary across schools).

In a further model (Model 15, Table A5), we also added prior attainment (in the form of z scores based on participants' average CAT scores or average Key Stage 2 SAT scores). This significantly improved the overall model fit. The variance in intercepts between schools was no longer significant; in other words, there was no evidence that a significant portion of the overall variance in time 2 total self-efficacy scores was accounted for at the school level. There was a significant effect of both time 1 score, $F(1, 542.15) = 227.64$, $p < .001$, and of prior attainment, $F(1, 253.67) = 43.66$, $p < .001$, on time 2 self-efficacy outcomes. Furthermore, after controlling for prior attainment, there was now a significant effect on time 2 scores of being in the Strategies group compared to the Texts group, $F(1, 23.81) = 10.58$, $p = .003$. There was, however, still no significant effect of being in the Phonics group compared to the Texts group, $F(1, 23.58) = 0.70$, $p = .412$.

Table A4: Models of the effects of the interventions on time 2 total self-efficacy outcomes

fixed effects	Model 13			Model 14			Model 15		
	b	SE	95% CI	b	SE	95% CI	b	SE	95% CI
t1 total self-efficacy score	0.56***	0.03	0.49, 0.62	0.54***	0.03	0.47, 0.60	0.55***	0.04	0.48, 0.62
Strategies group	0.68*	0.27	0.15, 1.21	0.70	0.40	-0.10, 1.50	1.20**	0.37	0.44, 1.96
Phonics group	.05	0.55	-0.47, 0.57	0.09	0.39	-0.70, 0.87	0.29	0.35	-0.43, 1.02
Prior attainment (z score)							0.96***	0.14	0.67, 1.24
random effects				σ^2	SE	95% CI	σ^2	SE	95% CI
school				.49***	0.21	0.21, 1.14	0.18	0.16	0.03, 1.07
model fit									
-2LL	3539.79			3525.127			2675.285		
$\Delta \chi^2(p)$				14.66***			849.84***		

* $p < .05$; ** $p < .01$; *** $p < .001$