The effects of phonics instruction on L2 phonological decoding and vocabulary learning: an experimental study of Chinese EFL learners

LI Sha, Zhejiang University

Robert Woore, University of Oxford Department of Education

Abstract

Phonological decoding – defined here as the ability to convert the written forms of an alphabetic writing system into the phonological forms they represent, using knowledge of the language's symbol-sound correspondences (SSC) – may play an important role in various aspects of second language (L2) learning, including the crucial task of vocabulary learning. Yet, many L2 learners appear to have low phonological decoding proficiency, even after extensive exposure to the language. Phonics instruction – explicit teaching of a language's SSC – has been found to improve phonological decoding amongst first language (L1) beginner readers. However, there is limited evidence concerning the effects of phonics instruction in L2. Addressing this gap, we conducted a quasi-experimental study in which two classes of first-year English majors (n=71) in two Chinese universities received twelve weeks of English phonics instruction. A comparison group comprising parallel classes in the same universities (n=67) received twelve weeks of instruction in English phonology, but without phonics. Pre- and post-tests measured participants' phonological decoding and vocabulary memorisation; the latter was assessed by immediate recall tests and recognition tests of the new words' phonological and orthographic forms. The phonics group significantly outperformed the comparison group in both post-tests, with medium effect sizes.

Keywords: phonological decoding; phonics instruction; vocabulary learning; classroom intervention; EFL

/ English as a Foreign Language

Introduction

Background

Over recent decades, the teaching of phonics – that is, explicit instruction in the systematic associations between written symbols and the sounds they represent in an alphabetic writing system – has become highly prominent in the teaching of early literacy in English as a first language (L1). Extensive research has shown the positive effects of phonics instruction on children's early reading and, to a lesser extent, spelling (Torgerson, Brooks & Hall, 2006). Similarly, official reviews of evidence in the US (NICHD, 2000), UK (Rose, 2006) and Australia (Rowe, 2005) have all concluded that systematic phonics instruction is the most effective way to teach young children to read.

It is perhaps no coincidence that phonics teaching has also gained popularity in second language (L2) contexts – including in China, the location of the current study. For example, Cao's (2017) analysis of the Chinese national research database (CNKI) found that the number of publications on phonics in EFL (English as a Foreign Language) teaching had risen steadily since the term was first mentioned in that database in 2005, reaching 27 publications in 2016. Huo & Wang (2017) also report that phonics now features in the official English curricula in many EFL countries and regions (p. 3), including Malaysia and Taiwan. In England, meanwhile, a prominent review of L2 pedagogy at secondary school level (Bauckham, 2016) recently recommended the 'direct and systematic teaching of phonics' in European languages (p. 12). However, in contrast to the considerable body of research underpinning phonics teaching in L1 (particularly English) settings, the evidence base in L2 contexts is much sparser

– particularly in ‘foreign’ language contexts, where the L2 is not a majority language of the wider social or educational setting. The current study aims to contribute to the small but growing number of studies in this area.

The importance of L2 phonological decoding

In L1 English, phonics instruction has been viewed mainly through the lens of reading. Phonics develops children’s proficiency in phonological decoding (or ‘recoding’) – that is, the ability to ‘convert the visual print into its corresponding spoken form’ (Nassaji, 2014, p.9) using knowledge of the systematic symbol-sound correspondences (SSC) in a given writing system. Phonological decoding allows children to ‘sound out’ words which they encounter in print for the first time, and thus ‘discover’ their meanings. For example, a child who has been taught the sounds of the letters <p>, <a> and <n>, on seeing the word ‘pan’ for the first time, can sound out these individual letters, blend the sounds together and recognize the resulting phonological form as the cooking utensil used at home. Underpinning this process is the child’s existing knowledge of the spoken language, to which she can associate the written forms she decodes.

‘Foreign’ language learners, by contrast, often encounter written forms at the very outset of learning the language; indeed, many of their first encounters with new words will be via their written forms. In other words, sounding out a word will often not help them discover its meaning, since they do not know the word orally in the first place. The role of phonological decoding in early reading comprehension is therefore likely to be more limited in many L2 contexts than in L1.

On the other hand, there are reasons to believe that phonological decoding may be a foundational skill, important for various other aspects of L2 learning. These include: motivation (Erler & Macaro, 2011); writing (Macaro, 2007); listening (via the ability to visualize the spoken forms that the learner hears: Field, 2004); and the crucial task of vocabulary learning. It is on this last issue that we will focus, since it forms a key strand of the current study.

Proficiency in phonological decoding is likely to support vocabulary learning, because it strengthens the links between the spoken and written forms of words, thus contributing to high quality lexical representations (Perfetti & Hart, 2002). Further, if a learner can decode an alphabetic language, then (to the extent that the language has regular and consistent SSC), he can generate for himself the spoken forms of unfamiliar written words he encounters. In other words, whilst he may not be able to 'discover' the meanings of new words by sounding them out, he can 'discover' their pronunciations, which will contribute to learning them. Recent empirical studies have indeed found evidence of a relationship between L2 decoding and intentional vocabulary learning (e.g. Hamada and Koda, 2008; Li, 2012). However, these studies are correlational and are thus unable to establish causality. Does increased decoding proficiency lead to improved word learning? To answer this question, experimental study designs are needed.

L2 decoding in the absence of phonics

Despite the likely importance of phonological decoding for wider L2 learning, many learners struggle to develop proficiency in this area in the absence of explicit instruction. This has been found in studies both of beginner, school-age learners (e.g. Erler, 2003, Erler & Macaro, 2011 and Woore, 2009 in the

UK) and of university students at higher L2 proficiency levels (e.g. Li, 2019 in China; Alghamdi, 2020 in Saudi Arabia). Additionally, a consistent finding in these studies is that learners' L2 decoding errors often arise through negative 'transfer' (Odlin, 1989) from their L1 writing system and L1 phonology. For example, many English school children decode French words using English SSC; many Chinese students decode English words using the SSC of *pinyin* (the Roman alphabetic representation of Chinese). This can be understood as the automatic triggering of L1-based associations by L2 input, where the L1 and L2 share the same alphabet[1].

In such cases, therefore, learners must override their L1-tuned decoding mechanisms, suppressing automatic L1-based symbol-sound associations where these differ from the L2. This will then allow them to establish new, L2-based associations. A useful theoretical lens through which to view this process is Schneider and Shiffrin's (1977) model of controlled versus automatic processing. Whereas automatic processing is rapid and involuntary, controlled processing is slow and conscious; it is what learners would use to 'work out' the pronunciations of L2 words by drawing on explicit knowledge of L2 SSC. Accordingly, Erler (2003) observed that some of her participants who took longer to read French words aloud actually pronounced them more accurately.

[1] As noted by one of our anonymous reviewers, errors in L2 decoding may also arise through the influence of the learner's L1 phonological system. An example would be where two different L2 phonemes are not distinguished in L1, as in the case of Japanese EFL learners who decode 'rock' as 'lock', because the phoneme /r/ does not occur in their L1 (Ota et al., 2009). Other L2 decoding errors may be due to constraints on syllable structure in L1. For example, Chinese EFL learners may decode the English pseudoword 'knaf' as /kənæf/ (with an epenthetic schwa), because syllable-initial consonant clusters are not permitted in their L1 (Li, 2019). However, such errors are not the main focus of this article.

Empirical studies of L2 phonics

L2 phonics instruction aims to address the kinds of problems outlined above. However, as noted above, to date there is only limited empirical evidence for its effectiveness. In particular, to our knowledge, there have been few experimental or quasi-experimental studies evaluating the effects of L2 phonics instruction in so-called ‘foreign language’ settings. As we do not have space to review all of the existing studies in this area, we will focus on those which we consider most robust and most relevant to the current study.

Huo and Wang (2017) synthesized the available evidence on the effectiveness of ‘phonological-based instruction’ (including phonics and phonological awareness) for young EFL learners from Kindergarten to Grade 6. They identified 15 studies published between 2000 and 2016 which met their inclusion criteria. However, most had significant methodological flaws: only three were considered sufficiently high quality to allow causal inferences to be drawn. Further, only one of these more rigorous studies (Dixon et al., 2011) looked specifically at phonics. Overall, Huo and Wang’s (2017) review found that phonological instruction was effective in developing ‘reading underlying skills’ (such as phonological decoding) with a moderate effect size; however, the effects on real word reading and reading comprehension were smaller, inconsistent, and in some studies, non-existent. The authors explain this on the basis that proficient decoding is not in itself sufficient for reading comprehension: rather, as argued above, a bank of oral language knowledge is also needed. However, a limitation of Huo and Wang’s (2017) review is that it focuses only on literacy development, and does not consider wider L2 outcomes, such as vocabulary acquisition. Further, it covers only young L2 learners, in contrast to the current study’s sample of university students.

At secondary school level, Coates et al. (2017) gave 24 Italian students (aged 17-18) ten weeks of L2 English phonics instruction (2 hours per week) in addition to their usual curriculum. A control group of 14 students received an equivalent amount of extra English tuition, but no phonics. The phonics group outperformed the control group on tests of English spelling, phonological decoding and productive oral language skills. However, the authors note that teacher effects could have influenced their findings.

Another recent study in secondary school L2 classrooms was Woore et al. (2018). In this 'cluster' RCT, 36 schools were allocated to one of three groups. All three groups read a series of challenging texts; one group additionally received phonics instruction, and another received reading strategy instruction. At baseline, the groups performed equivalently in phonological decoding, reading comprehension and vocabulary knowledge. At post-test, all three groups had made significant progress in reading comprehension, with no evidence of an advantage for any one group over the others. There was, however, some evidence of an advantage for the phonics group in phonological decoding, and for the phonics and strategies groups (most clearly the phonics group) in vocabulary development. However, any longer-term impact of the instructional programmes was unclear, due to considerable attrition between the immediate post-tests and delayed post-tests six months later.

Finally, at university level, Sturm (2013) evaluated a programme of 'explicit phonetics instruction', including some phonics, in L2 French. Eleven undergraduate students in the US received the phonetics instruction; eleven others, who followed various other French courses, formed a control group. On a post-intervention reading aloud test, the phonetics group were significantly more accurate than the control group, with a large effect size. However, the study suffers from several limitations. For example, participants were not randomly allocated to conditions; indeed, the phonetics group had elected to follow this course, perhaps indicating an existing interest in improving their French pronunciation. Further, the author herself taught the experimental group but not the control group, and testing was apparently not blind to condition. Further studies at university level are therefore needed to test out these findings, and to extend them to different L1/L2 combinations.

Summary and research questions

To summarize, previous research provides some evidence that L2 phonics instruction may help learners to become more accurate phonological decoders; without such instruction, they may struggle to become proficient in this area, and find it difficult to overcome entrenched L1-based symbol-sound associations (where the L1 and L2 writing systems share the same alphabet). However, in contrast to early L1 literacy instruction, the biggest effects of L2 phonics instruction may not be on reading comprehension, but rather on other aspects of learning the language, particularly vocabulary learning. Yet, the evidence base on the effectiveness of L2 phonics instruction remains limited. Against this backdrop, the current study addresses the following two research questions:

For a sample of Chinese university EFL learners, does a programme of systematic English phonics instruction lead to:

(1) improved phonological decoding of unfamiliar English words?

(2) improved intentional learning of unfamiliar English words?

Method

Participants

Participants in this study were four intact classes of first-year English majors (n=138) studying in two universities in Wuhan, China (henceforth Universities A and B)[2]. All participants were L1 Chinese speakers who had received Chinese-medium education throughout high school. None had family members who were native speakers of English, as revealed by a questionnaire conducted before the main data collection.

Study design

Individual, random assignment of participants to experimental conditions would have been desirable, but was not possible due to practical constraints. A quasi-experimental design was therefore adopted, in which one intact class from each university was randomly assigned to the intervention group, and the other to the comparison group (Table 1). Given the small, clustered nature of the sample, we conceptualize our study as a 'proof of concept' trial.

Table 1. Numbers of participants in each university and in each condition

	Intervention	Comparison	Total
University A	24	21	45
University B	47	46	93
Total	71	67	138

[2] Data was originally gathered from a third university, but was discarded because the participants there were concurrently learning French, which was found to interfere with the English phonics instruction. Further details of these participants and their outcomes can be found in Li (2019).

Two tests were administered to check that the intervention and comparison groups were well matched: an English proficiency test (National College Entrance English Exam, NCEEE); and an English vocabulary test (British Picture Vocabulary Scale, BPVS). Two one-way ANOVAs found no significant differences between the groups in terms of either NCEEE scores ($F(1, 136) = 1.76$, $p = .19$) or BPVS scores ($F(1, 136) = .03$, $p = .86$) (Table 2).

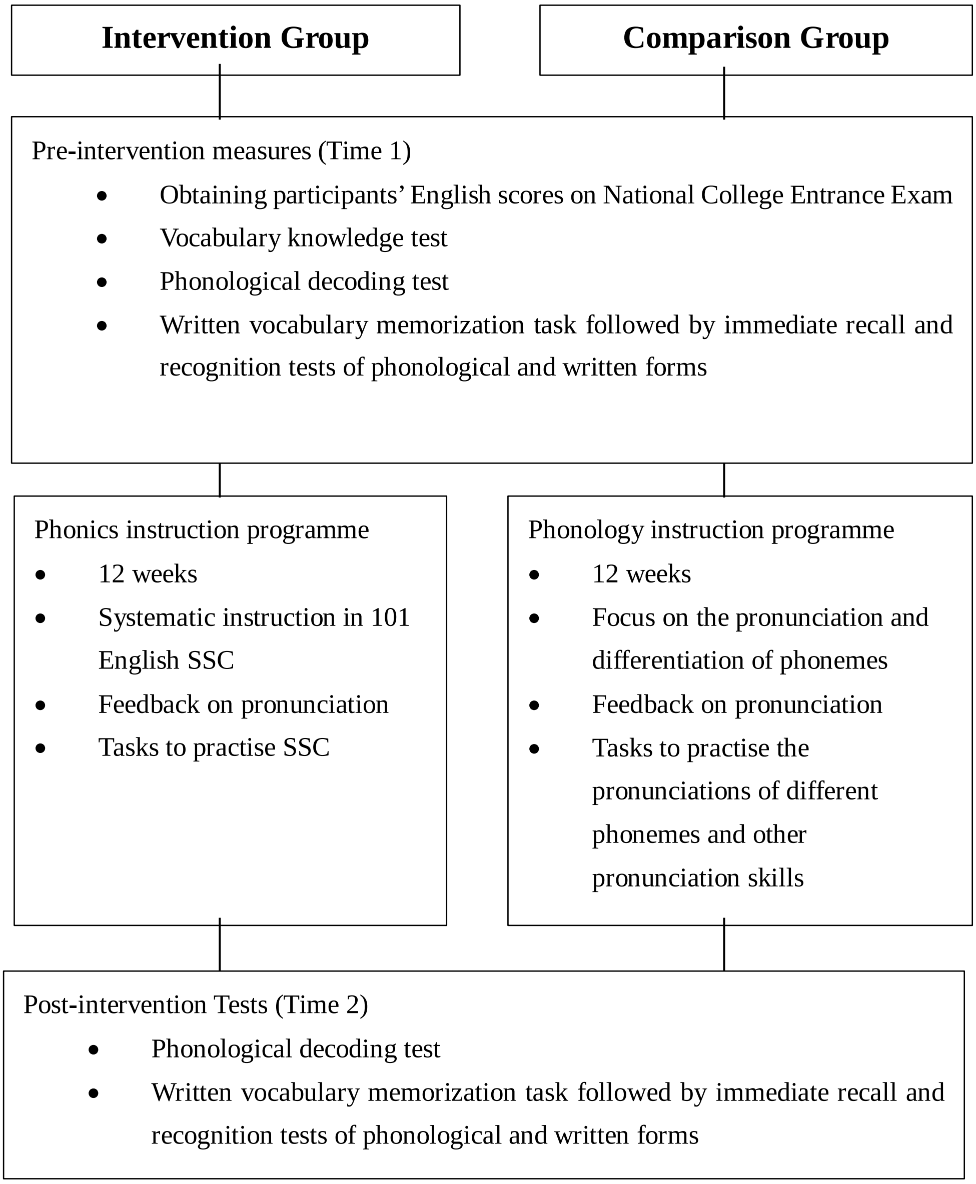
Table 2. Participants' scores on the NCEEE and BPVS

	NCEEE (out of 150)		BPVS (out of 70)	
	Mean	SD	Mean	SD
Intervention (N=71)	125.1	7.3	56.4	7.7
Comparison (N=67)	123.4	7.7	56.6	5.6

Pre- and post-tests, administered immediately before (t1) and immediately after (t2) the intervention, comprised: (a) an English phonological decoding test; and (b) a vocabulary memorisation task (based on the written forms of the words), immediately followed by recall and recognition tests of both the spoken and written forms of the new words. All tests were administered individually by the first author in a quiet, unoccupied classroom. Delayed post-tests, measuring the longer-term impact of the intervention, were unfortunately impracticable due to resource constraints.

As noted by one of our anonymous reviewers, ideally, we would also have included measures of working memory and phonemic awareness in our study, to check that the groups were comparable on these variables. However, the testing burden for participants was already high, and we were concerned not to increase this further, lest we affect recruitment or retention. Further, in an earlier study (Li, 2012) exploring the correlation between L2 English decoding proficiency and vocabulary learning amongst a very similar sample of Chinese university EFL learners, we found very little variation in participants' working memory scores, as measured by a forward digit span test (Hoppe et al., 2000). In terms of phonemic awareness, we considered that this variable was already adequately controlled by our baseline tests, given that phonemic awareness is likely to correlate with (and potentially contribute to) phonological decoding. We therefore felt justified in omitting working memory and phonemic awareness measures from our study, although we acknowledge their absence as a limitation.

Figure 1. Overview of the research design



Intervention design

Both the phonics instruction and the phonology instruction took place in participants' phonology class, which was a mandatory course for all first-year English majors. The intervention group received English phonics instruction, while the comparison group received their usual programme of instruction in English phonology, without phonics. (Ideally, an additional 'true control' group would have been included, receiving neither phonics nor phonology instruction; however, this was not possible within the constraints of the host universities.) Both programmes of instruction comprised 90 minutes of teaching per week for twelve weeks. Both were taught by the first author, reducing the risk of teacher effects. However, the possibility of implicit bias towards one or other of the groups must be acknowledged. To mitigate this risk, and to ensure help ensure fidelity to condition, all sessions (in both groups) were observed by a teacher of phonology in the respective universities. Nonetheless, we acknowledge that having an independent person, not one of the authors, teach both the phonics and phonology courses would have been preferable.

The phonics programme used the textbook *Read Write Inc.: Fresh Start* (Miskin, 2011), originally designed for struggling readers of English in the UK, aged 9-13 years. In total, 101 graphemes (corresponding to 44 phonemes) were taught, along with their most frequent pronunciations, as well as 27 common word endings (e.g. *–tion*, *-al*) (Appendix 1). The teaching of the target SSC was supported by use of the International Phonetic Alphabet (IPA, 2005). Typically, each class began by reviewing SSC covered previously, before new SSC were introduced and modelled for students. There were then multiple opportunities to practise these SSC in whole-class and pair work, such as reading words aloud containing the new SSC with corrective feedback from the teacher.

For the comparison group, the textbook was *English Pronunciations and Intonations for Communication* (Wang, 2005), one of the most popular phonology textbooks for English majors in China. The programme covered (a) the basic concepts of syllables, stress and rhythm; (b) systematic instruction in the pronunciation of 44 phonemes and their corresponding symbols in the International Phonetic Alphabet; (c) pronunciation skills, such as intonation and stress patterns (Appendix 2). A typical class began with a review of material covered previously, followed by the presentation of new phonemes with modelling by the teacher. Participants then practised what they had learnt in reading, listening and speaking tasks, conducted as a whole class or in pairs, with corrective feedback from the teacher. We acknowledge that the materials provided to the comparison group did provide opportunities for incidental learning of English SSC (because words exemplifying a given sound were encountered in both spoken and written forms concurrently). However, in contrast to the intervention group, no explicit instruction in English SSC was provided. Thus, the key difference between the groups was the presence or absence of English phonics instruction. Appendix 3 contains a more detailed overview of both the intervention and comparison programmes of instruction.

Data collection

At t1 and t2, participants' English decoding proficiency was measured using a section of the *Woodcock Reading Mastery Tests* (WRMT: Woodcock, 2011), where participants read aloud 28 pseudowords of increasing difficulty (appendix 4), such as 'op', 'tayed', 'throbe', 'whumb' and 'monglustamer'. Two test forms, A and B, were used. A pilot study with 20 students found no significant difference between the mean scores achieved in these forms, as revealed by a t-test ($t(19) = -.74$, $p = .47$), suggesting that they were of similar difficulty. A split-block counterbalanced design was adopted for the decoding tests: half of the participants in each class took Form A at t1 and Form B at t2, and vice versa for the other half of participants. Participants read the pseudowords at their own pace. This differed from the method used for the same test in many other studies, where the items appeared for a fixed period of time (usually a few seconds: e.g. Hamada & Koda, 2008, 2010). The reason for our decision relates to the possible role, noted above, of controlled (and therefore slower) processing in overcoming automatic L1-based symbol-sound associations. Imposing a time limit might have constrained participants' controlled processing and thus their ability to deploy any explicit knowledge of English SSC gained through the instruction programmes. This, in turn, may have prevented them from reaching maximal accuracy in reading aloud. Nonetheless, the total time taken for each participant to complete the decoding test was recorded (see below), as a measure of decoding speed.

At t1 and t2, participants also completed a vocabulary learning task comprising 10 real, low frequency English words (appendix 5), such as 'argent', 'ploidy', 'mayhap', 'turgor' and 'rheumy'. Similar to the decoding test, two different sets of words (A and B) were used in a split-block counterbalanced design. A two-tailed t-test found no significant difference in the mean bigram frequency of the words in the two test forms ($t(9) = -.69$, $p = .52$), suggesting that they were well matched orthographically. Participants saw each English word, together with its Chinese character translation, for 20 seconds (indicated by piloting to be an appropriate length of time) before advancing automatically to the next item. They were told to learn the spoken and written forms of the words, and their meanings, and that they would be tested on these afterwards. Hence, this was an 'intentional' rather than 'incidental' learning task, according to Hulstijn's classification (2001).

Vocabulary knowledge is a multidimensional construct (Nation, 2001). It was of interest to examine whether phonics instruction contributed differentially to different aspects of vocabulary learning. Thus, participants completed four tests immediately after the memorization task:

(a) Oral recall: they saw the Chinese translations and said the English words;

(b) Written recall: they saw the Chinese translations and wrote the English words;

(c) Aural recognition: they listened to audio recordings of the English words and wrote Chinese translations;

(d) Written recognition: they saw the English words and wrote Chinese translations.

The ordering of the four tests (as listed above) was designed to reduce the impact of practice effects (e.g. seeing the correct spelling in the written recognition test could have primed participants for the written recall test if the ordering had been different).

Analyses

The scoring of all tests was completed blind to condition by the first author. For the decoding test, the pronunciation of each pseudoword was marked either right (1 point) or wrong (0 points). To check reliability of the scoring, 20% of the pronunciations at each time point were additionally marked by a Chinese teacher from University A. Interrater agreement of 100% was found at t1 and 99.8% at t2.

Many previous studies involving word naming tasks have measured ‘naming latencies’, that is, the time between the appearance of each stimulus and the onset of participants’ pronunciation of it (e.g. Hamada & Koda, 2008). However, this approach overlooks the fact that many participants may hesitate, make false starts (e.g. /mɒ - mɒn - mɒnglʌstəmə/ in our data) and/or self-correct their initial pronunciations (e.g. /mɒnglestəmə - mɒnglʌstəmə/) – particularly in an L2, given the potentially greater role for controlled rather than automatic processing. Therefore, we took a different approach, measuring the total time taken to decode all 28 test items (from appearance of the first stimulus to the end of the pronunciation of the last item). Whilst this is a less precise measure of decoding speed, we considered it sufficiently accurate for our purposes. It also has the advantage of including the various hesitation and self-correction phenomena described above.

For the vocabulary recall and recognition tests, one point was given for each correct response and zero for each incorrect response. In the recall tests, spelling and pronunciation had to be fully correct (allowing for some degree of foreign accent in the latter case) to score the point. To check reliability of scoring, a 20% subsample of test results at each time point were additionally marked by the same teacher from University A. Interrater agreement of 100% was found for the scoring of all tests.

Results

RQ1. Phonological decoding

Decoding accuracy scores

Table 3 shows the number of pseudowords pronounced correctly in the decoding test. After checking that the assumptions of Normality and homogeneity of variances between the groups were met at both t1 and t2, a two-way mixed factorial ANOVA was conducted with one within-subjects variable (time) and one between-subjects variable (condition). This revealed a significant main effect of time with a large effect size, $F(1, 136) = 72.74$, $p < .001$, $r = .59$, as well as a significant main effect of condition with a small effect size, $F(1, 136) = 7.89$, $p < .01$, $r = .23$. The interaction between time and condition was also significant, $F(1, 136) = 34.80$, $p < .001$, with a medium effect size ($r = .45$). Thus, the phonics group made significantly greater gains in accuracy of English decoding than the comparison group.

Table 3. Number of words decoded accurately out of 28

	t1		***t2***	
Group	Median	Mean (SD)	Median	Mean (SD)

Intervention	11	10.6 (4.0)	16.0	15.1 (3.8)
Comparison	11	10.0(4.1)	12.0	11.5 (4.3)

Decoding speed

Table 4 shows the total time spent on the decoding test. For this variable, the assumptions of Normality and homogeneity of variances were violated at t2. Therefore, non-parametric tests were used. A Wilcoxon Signed Ranks test found that, for both the intervention and comparison groups, overall decoding time differed significantly between t1 and t2: intervention group: $Z = -3.02$, $p < .01$, $r = -.26$; control group: $Z = -6.95$, $p < .001$, $r = -.58$. Since the mean decoding time was greater at t1 than t2 for both groups, this shows that both groups were significantly faster decoders at t2. Second, Mann-Whitney tests found that the decoding times of the two groups did not differ significantly at t1 ($U = 1912$, $Z = -1.95$, $p = .06$) but did differ significantly at t2, with a medium effect size ($U = 569.5$, $Z = -7.71$, $p < .001$, $r = -.46$). Thus, over the period of the study, the comparison group showed a significantly greater decrease in decoding time (i.e. they showed a significantly greater increase in speed) compared to the intervention group. Table 4 also shows a striking reduction in the standard deviation of the decoding time for the comparison group: this is about 2.5 times lower at t2 than t1. By contrast, the standard deviation of the intervention group remains similar.

Table 4. Total time spent on the decoding test, to the nearest second

	t1		***t2***	
Group	Median	Mean (SD)	Median	Mean (SD)
Intervention	75	75 (19)	66	69 (17)
Comparison	69	70 (17)	47	47 (7)

RQ2. Vocabulary memorisation test results

This section reports the results of the four vocabulary measures, in the order in which they were completed by participants: oral recall, written recall, aural recognition, written recognition. Outcomes on these measures are summarized in Table 5, and statistical analyses are presented below. Since the assumption of Normality was violated for all four variables at both time points, non-parametric tests were used.

Table 5. Groups' scores on the vocabulary memorization tests (All tests are out of 10)

		t1		t2	
Variable	**Group**	**Median**	**Mean (SD)**	**Median**	**Mean (SD)**
Oral recall	**Intervention**	1	1.0 (0.9)	5	4.2 (1.5)
	Comparison	1	1.0 (0.9)	2	2.5 (1.4)
Written recall	**Intervention**	3	3.1 (1.6)	6	6.1 (2.1)
	Comparison	3	3.0 (1.6)	5	5.1 (2.3)
Aural recognition	**Intervention**	6.0	5.6 (2.4)	9.0	8.4 (2.8)
	Comparison	6.0	5.7 (2.3)	7.0	7.0 (2.0)
Written recognition	**Intervention**	8.00	7.1 (2.1)	10.00	9.1 (1.1)
	Comparison	8.00	7.2 (2.2)	10.00	9.0 (1.5)

Oral recall test

On the oral recall test, a Wilcoxon Signed Ranks test found a significant difference between the t1 and t2 scores for the intervention group (Z = -6.59, p < .001, r = -.55) but not for the comparison group (Z = -1.70, p = .09). As the intervention group's mean score was higher at t2 than t1, this indicates that this group (but not the comparison group) made significant gains in the ability to recall newly-learnt oral word forms over the period of the study. Second, Mann-Whitney tests found that the scores of the two groups were not significantly different at t1 (*U* = 2323, *Z* = -.25, *p* = .80) but were significantly different at t2, with a medium effect size (*U* = 880.5, *Z* = - 6.52, *p* < .001, *r* = .39). This suggests that the phonics instruction led to significantly greater progress than the phonology instruction in participants' ability to recall the oral forms of newly-learnt words.

Written recall test

On the written recall test, a Wilcoxon Signed Ranks test found significant differences between the t1 and t2 scores for both the intervention group (*Z* = -6.90, *p* < .001, *r* = -.22) and the comparison group (*Z* = -5.72, *p* < .001, *r*= -.21). Since mean written recall scores were higher at t2 than t1 for both groups, this indicates that both groups made significant gains in their ability to recall newly-learnt written word forms. Second, Mann-Whitney tests found that the scores of the two groups were not significantly different at t1 (*U* = 2372.5, *Z* = -.03, *p* = .98) but were significantly different at t2, with a small effect size in favour of the phonics group (*U* =1771.5, *Z* = -2.65, *p* < .01, *r* = -.22). This again indicates that the phonics instruction led to significantly greater gains than the phonology instruction in participants' ability to recall the written forms of newly-learnt words.

Aural recognition test

On the aural recognition test, a Wilcoxon Signed Ranks test found significant differences between t1 and t2 scores for both the intervention group (*Z* = -6.58, *p* < .001, *r* = -.57) and the comparison group (*Z* = -4.38, *p* < .001, *r* = -.37). Since both groups' aural recognition scores were higher at t2 than t1, this indicates that both groups made significant gains over the period of the study. Next, Mann-Whitney tests found that the scores of the two groups were not significantly different at t1 (*U* = 2347, *Z* = -.14, *p* = .89) but were significantly different at t2, with a small effect size in favour of the phonics group (*U* =1077, *Z* = -5.65, *p* < .001, *r* = -.34). Thus, the phonics instruction led to significantly greater progress than the phonology instruction in participants' ability to recognize the spoken forms of newly-learnt words.

Written recognition test

Finally, on the written recognition test, a Wilcoxon Signed Ranks test found significant differences between t1 and t2 scores for both the intervention group (*Z* = -6.52, *p* < .001, *r* = -.56) and the comparison group (*Z* = -6.01, *p* < .001, *r* = -.50). Since both groups' written recognition scores were higher at t2 than t1, this indicates that both groups made significant gains in this area over the period of the study. Next, Mann-Whitney tests found that the scores of the two groups were not significantly different at t1 (*U* = 2292, *Z* = -.37, *p* = .71) and nor were they significantly different at t2 (*U* = 2313.5, *Z* = -.15, *p* = .88). This shows that the phonics instruction did not lead to significantly greater progress than the phonology instruction in participants' ability to recognize the written forms of newly-learnt words. This contrasts with the findings from the other three tests relevant to Research Question 2: the written recognition test was the only one of the four tests where the intervention group did not make significantly greater gains than the comparison group.

Discussion

Phonological decoding test results

Addressing research question 1, participants who received systematic phonics instruction made significantly greater progress in English decoding accuracy than those in the comparison group, with a medium effect size. On average, at t1, both groups performed equivalently; but at t2, the phonics group decoded 4 additional pseudowords correctly out of 28 (an improvement of 14 percentage points). These findings provide evidence for the effectiveness of phonics instruction in promoting accuracy of phonological decoding in L2, based on a sample of university-level learners of English with Chinese L1. Our study thus contributes to the small but growing body of evidence for the effectiveness of phonics instruction on L2 decoding at various educational levels, including primary school (e.g. Huo & Wang, 2017), secondary school (e.g. Coates et al., 2017; Woore et al., 2018) and university (e.g. Sturm, 2013).

It is also worth noting that, whilst the comparison group in our study made significantly less progress in English decoding than the intervention group, they did nonetheless achieve significantly higher decoding scores at t2 than t1. This was unexpected: previous studies have found that many L2 learners, in the absence of phonics instruction, make very limited progress (or even none at all) in phonological decoding, even over extended periods of time, at least in contexts where learners' L1 and L2 are both alphabetic languages (e.g. Woore, 2009). The different finding in the current study may reflect the fact that the instruction received by the comparison group was designed to raise participants' English phonological awareness. For example, they were taught the concept of syllables and practised intra-word analysis via word segmentation and syllable counting. Studies of L1 English reading have consistently demonstrated that phonological awareness is a key predictor of literacy development (e.g. August et al., 2014). Thus, these activities may have helped comparison participants to increase their familiarity with English phonological structures, thereby laying foundations for improved decoding.

Proficient decoding comprises not only accuracy, but also speed. In our study, we used a relatively crude measure of decoding speed, but this nonetheless yielded interesting findings. Between t1 and t2, both groups showed a significant decrease in total decoding time (perhaps partly as a result of a practice effect). However, the decrease was significantly greater for the comparison group (medium-large effect size) than for the intervention group (small effect size). In other words, by the end of the study, the intervention group had become more accurate in their English phonological decoding, but the comparison group had become faster. This ran counter to our initial expectations: we had assumed that the phonics instruction would lead to an advantage in both accuracy and speed of decoding.

One possible explanation may be that the phonics instruction developed participants' *explicit* knowledge of English SSC in the first instance. Since the decoding test was not speeded, intervention participants had time to think about their responses, allowing them to use controlled processing (Schneider & Shiffrin, 1977) to consciously override any inappropriate, L1-based (i.e. *pinyin*-based) symbol-sound associations which may have been automatically triggered by the L2 input. They could then draw instead on any explicit knowledge of L2 SSC gained from the phonics teaching. As a result, their pronunciations may have been more accurate, but at a cost of greater processing time. The comparison group, by contrast, may have been more likely to continue relying on L1 (*pinyin*) SSC when pronouncing the English pseudowords, resulting in pronunciations which were faster but less accurate than those of the intervention group.

It is also interesting that the intervention group showed greater variance than the comparison group in their decoding speed at t2: the standard deviation in t2 decoding times was almost 2.5 times lower for the comparison group. This observation is again consistent with the view that intervention participants were making greater use of controlled processing than the comparison group. Such controlled processing would in turn be associated with greater individual differences (e.g. because of differing levels of motivation or knowledge), compared to the more automatic application of L1-based SSC. In the longer term, it would be hoped that the phonics participants would go on to automatize their more accurate decoding through practice. To test this hypothesis, delayed post-tests would be needed, which were however beyond the scope of the current study.

In support of the above interpretation of our findings, further examination of the data revealed that some intervention participants who took longer than others to complete the decoding test also produced more accurate pronunciations of the test items. This echoes similar findings reported by Erler (2003) and Li (2012). One implication of this is that, in any future research aiming to measure the effects of L2 phonics instruction, a non-speeded (or at least, generously timed) test may be best able to capture changes in the accuracy of participants' L2 decoding, at least in the short term.

Vocabulary memorisation test results

Research question 2 investigated the effects of phonics instruction on intentional word learning. Previous studies have reported a positive relationship between L2 phonological decoding and the learning of new orthographic and phonological word forms (e.g. Hamada & Koda, 2008; Li, 2012). There are reasons to believe that this relationship may be a causal one: i.e. that better L2 decoding proficiency facilitates the acquisition of L2 vocabulary, particularly in instructed contexts, where many new words are encountered first in written form. However, only one study (to our knowledge) has thus far attempted to find empirical evidence of this causal relationship (Woore et al., 2018). That study found tentative evidence that participants who had received phonics instruction acquired more vocabulary, over the period of the study, than comparison groups.

Similarly, in the current study, after an intentional learning task involving ten low frequency English words, intervention participants showed a significant advantage over the comparison group, at t2, in their ability to: recall the spoken forms of the new words (with a medium effect size); recognize the words' spoken forms (small effect size); and recall their written forms (small effect size). On average, compared to the comparison group, intervention participants correctly recalled and recognized the spoken forms of around 1.5 additional words out of 10, and correctly recalled the written form of around 1 additional word out of 10. We believe that such an advantage (amounting to 10-15%) on a single vocabulary learning task is substantial, given the central importance of vocabulary learning for the development of all four language skills (Milton, 2013) and the large number of words to be learnt.

For the two testing modes involving the phonological forms of the new words (spoken recall and spoken recognition), it is clear why the ability to decode the written forms (in which the new words were presented) conferred an advantage in terms of storing accurate phonological forms. For the written recall task, the phonics instruction may have helped participants to establish stronger and more reliable representations of the novel orthographic forms. This perhaps resulted from more detailed intra-word analysis and the generation of more accurate phonological forms, which could then reinforce participants' orthographic representations in long-term memory. By contrast, the two groups performed equivalently on their ability to recognize the written forms of the new words. This is unsurprising: of the four vocabulary tests, written word recognition is intuitively the least likely to implicate phonological decoding. Further, both groups of participants already achieved high scores in the written recognition test at t1 (mean score >7/10): therefore, this was something in which they were already skilled, leaving less room for improvement over the course of our study.

Finally, we would note the value, when evaluating L2 phonics instruction, of measuring the development of vocabulary knowledge in a range of different ways, as we did in the current study. Written word forms may be the most convenient basis for any recall and recognition tests, but by including the spoken forms as well, we were able to gain a clearer and more nuanced picture of the impact of the phonics instruction on vocabulary learning.

Limitations

The current study, which was small in scale and conducted with limited resources, inevitably has several limitations. First, whilst we avoided the problem of teacher effects (as seen in phonics experiments such as Coates et al., 2017 and Sturm, 2013) by ensuring that both the intervention and comparison groups were taught by the same person, we were unable to avoid the attendant problem that this teacher – who was also the first author – may have been biased, even if only implicitly, towards the phonics condition. Second, allocation to the experimental and comparison groups was done at the level of intact classes rather than individual participants, introducing the possibility that pre-existing differences between the groups might confound the results. A third, associated problem is that we had only two groups per arm of the experiment. To mitigate these problems, however, we ensured that from each participating university, one class was allocated to the experimental group and one to the comparison group (rather than having one university per condition). Further, we ensured that the groups were matched in terms of English proficiency and vocabulary knowledge. Although we would ideally also have demonstrated their equivalence in terms of working memory capacity and phonemic awareness, we do not expect the groups to have differed on these variables. Fourth, we would ideally have included a third, 'true control' group who received neither phonics nor phonology instruction, but this was not possible given prevailing constraints.

Conclusions and implications

Phonics instruction is well established in the teaching of early literacy in L1 contexts and appears to be gaining popularity in L2 contexts too. The current study added to the small but growing body of research examining the effects of L2 phonics instruction for so-called 'foreign language' learners, using a quasi-experimental, pre-/post-test design and a sample of Chinese university students majoring in English. The findings corroborate those of previous studies: namely, a systematic, twelve-week programme of L2 phonics instruction led to improvements in the accuracy of participants' phonological decoding, with a medium effect size.

In L1 settings, phonics is generally viewed through the lens of learning to read (and, to a lesser extent, spell). However, we believe that its key benefits in L2 learning may lie elsewhere than reading. In the current study, we looked in particular at the effects of phonics instruction on vocabulary learning. Following an intentional word learning task, participants who had received phonics instruction significantly outperformed the comparison group in recalling the words' spoken forms, recognizing their spoken forms and recalling their written forms. There was, however, no advantage for either group on recognition of the words' written forms. Again, this echoes the findings of previous research – though, to our knowledge, only one other study has investigated the effect of L2 phonics teaching on vocabulary knowledge (Woore et al., 2018). Given the central importance of vocabulary acquisition in all aspects of L2 learning and use, this finding potentially has important pedagogical implications. Clearly, however, further studies – larger in scale, and addressing some of the methodological limitations of the current one – are needed before pedagogical recommendations can be made.

The current study used a sample of university-level EFL learners with considerable prior exposure to the language. Given the potential benefits of phonics instruction outlined above, it would be worthwhile to explore its effects on younger L2 learners, in order to see whether vocabulary learning could be enhanced at earlier stages of English learning. In future studies, additional outcome measures would also be desirable beyond vocabulary learning, including (but not limited to) reading comprehension, listening comprehension and motivation for learning the language.

References

Alghamdi, H. (2020). L2 Decoding and reading comprehension in Saudi foundation year ESL students: An investigation into readers' decoding abilities, outputs, experiences, and processes, and the contribution of their L2 decoding proficiency to their L2 reading comprehension. Unpublished doctoral thesis, University of Oxford.

August, D., McCardle, P., & Shanahan, T. (2014). Developing literacy in English language learners: Findings from a review of the experimental research. School Psychology Review, 43, 490-498.

Bauckham, I. (2016). Modern foreign languages pedagogy review: A review of modern foreign languages teaching practice in key stage 3 and key stage 4. Review report written for the Teaching Schools Council. Retrieved from https://www.tscouncil.org.uk/wp-content/uploads/2016/12/MFL-Pedagogy-ReviewReport-2.pdf (Accessed 03 June 2021).

Cao, M. (2017). An analysis of phonics teaching in mainland China. *Journal of Language Teaching and Research*, *8*, 286-290.

Coates, R. A. G., Gorham, J., & Nicholas, R. (2017). The Efficacy of Phonics-Based Instruction of English as a Second Language in an Italian High School: A Randomised Controlled Trial. *GIST Education and Learning Research Journal*, 15, 29-67.

Dixon, P., Schagen, I., & Seedhouse, P. (2011). The impact of an intervention on children's reading and spelling ability in low-income schools in India. *School Effectiveness and School Improvement*, *22*, 461-482.

Erler, L. (2003). Reading in a foreign language--- near-beginner adolescents' experience of French in English secondary schools. Unpublished doctoral thesis, University of Oxford.

Erler, L., & Macaro, E. (2011). Decoding ability in French as a foreign language and language learning motivation. *The Modern Language Journal*, *95*(4), 496-518

Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down?. *System*, 32, 363-377.

Hamada, M. and Koda, K. (2008). Influence of first language orthographic experience on second language decoding and word leaning. *Language Learning*, 58, 1-31.

Hamada, M. and Koda, K. (2010). The role of phonological decoding in second language word-meaning inference. *Applied Linguistics*, 31, 513-531.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 258–286). Cambridge: Cambridge University Press.

Huo, S., & Wang, S. The effectiveness of phonological-based instruction in English as a foreign language students at primary school level: A research synthesis. *Frontiers in Education*, 2, 1-13.

Li, S. (2012). Decoding ability and vocabulary acquisition in second language English- A study of Chinese advanced EFL learners. Unpublished master thesis, University of Oxford.

Li, S. (2019). *The effects of phonics instruction on L2 phonological decoding and vocabulary learning*. Unpublished doctoral thesis, University of Oxford.

Macaro, E. (2007). Language learner strategies: Adhering to a theoretical framework. *Language Learning Journal*, 35, 239-243.

Macaro, E., & Erler, L. (2008). Raising the achievement of young-beginner readers of French through strategy instruction. *Applied linguistics*, 29, 90-119.

Miskin, R. (2006). *Read Write Inc. Fresh Start: Handbook*. Oxford: Oxford University Press.

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), L2 vocabulary acquisition, knowledge and use. New perspectives on assessment and corpus analysis (pp. 57–78). EUROSLA Monographs Series 2.

Nassaji, H. (2014). The role and importance of lower-level processes in second language reading. *Language Teaching*, 47, 1–37.

Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network (2000). The relation of child care to cognitive and language development. *Child Development*, 71, 960-980.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition*, 111, 263-269.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, 11, 67-86.

Rose, J. (2006). *Independent review of the teaching of early reading: Final report*. London: Department for Education and Skills Publications.

Rowe, K.J. (Chair) (2005a). *Teaching reading literature review: A review of the evidence-based research literature on approaches to the teaching of literacy, particularly those that are effective in assisting students with reading difficultie*s. A report of the Committee for the National Inquiry into the Teaching of Literacy. Canberra, ACT: Australian Government Department of Education, Science and Training.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing. I. Detection, search, and attention. *Psychological Review*, 84, 1–66.

Sturm, J. L. (2013). Explicit phonetics instruction in L2 French: A global analysis of improvement. *System*, 41, 654-662.

Torgerson, C., Brooks, G., & Hall, J. (2006). *A systematic review of the research literature on the use of phonics in the teaching of reading and spelling*. Nottingham: DfES Publications.

Wang, G. (2005). *English Pronunciations and Intonations for Communication*. Beijing: Higher Education Press.

Woodcock, R. W. (2011). *Woodcock Reading Mastery Tests: WRMT-III*. Circle Pines: Pearson.

Woore, R. (2009). Beginners' progress in decoding L2 French: Some longitudinal evidence from English modern foreign languages classrooms. *Language Learning Journal*, 37, 3-18.

Woore, R., Graham, S., Porter, A., Courtney, L., & Savory, C. (2018). Foreign language education: Unlocking reading (FLEUR): A study into the teaching of reading to beginner learners of French in secondary school. Final report to the Nuffield Foundation. Retrieved from https://ora.ox.ac.uk/objects/uuid: 4b0cb239-72f0-49e4-8f32-3672625884f0 (Accessed 03 June 2021).

Appendix 1. SSC taught each week

Week number	Graphemes instructed
1	<b> <bb> (/b/); <p> <pp> (/p/); <t> <tt> <ed> (/t/); <d> <dd> <ed> (/d/); <k> (/k/); <g> <gg> (/g/); <h> (/h/); <s> (/s/); <m> <mm> (/m/); <r> <rr> (/r/); <f> <ff> (/f/); <l> <ll> (/l/);
2	<sh> (/ʃ/); <th> (/θ/, /ð/); <mb> (/m/); <wr> (/r/); <ss> (/s/); <ng> (/ŋ/) <nk> (/ŋk/); <ck> (/k/); <tch> (/tʃ/); <a> (/æ/); <e> (/e/); <i> (/ɪ/); <o> (/ɒ/); <u> (/ʌ/)
3	<a> (/æ/); <e> (/e/); <i> (/ɪ/); <o> (/ɒ/); <u> (/ʌ/); <ay> (/eɪ/); <ee> (/iː/); <igh> (/aɪ/); <c> <ck> (/k/)
4	<ow> (/əʊ/); <oo>(/u/); <ar> (/ɑː/); <th> (/θ/); <ch> (/tʃ/); <kn> <n> <nn> (/n/); <qu> (/kw/); <wh> <w> (/w/); <x> (/ks/); <y> (/j/)
5	<y> (/ɪ/); <or> <oor> <ore> (/ɔː/); <ch> (/tʃ/); <kn> (/n/); <se> (/s/); <ir> <ur> (/ɜː/)； <z> <zz> <s> (/z/)
6	<ou> (/aʊ/); <oy> <oi> (/ɒɪ/); <kn> (/n/); <se> (/s/); <bt> (/t/); <j> <dge> <ge> (/dʒ/); <wh> (/w/); <tch> (/tʃ/); <z> <zz> <s> <ze> (/z/)
7	<ay> <a-e> (/eɪ/); <ie> <i-e> (aɪ/); <ee> <ea> <ae> (/iː/); <ch> (/tʃ/); <se> <c> (/s/)
8	<ow> <o-e> (/əʊ/); <or> <aw> <au> (/ɔː/); <ear> <eer> (/ɪə/), <ce> <se> (/s/)
9	<air> <are> (/eə/); <ir> <ur> (/ɜː/); <ou> <ow> (/aʊ/); <sh> <ti> <ci> (/ʃ/);
10	<ai> <ey> <aigh> <eigh> (/eɪ/); <ow> <o> <oa> (/əʊ/); <oo> <u-e> <ue> <ew> (/uː/); <ire> (/aɪə/);
11	<-le> <-il> <-el> <-al> (/əl/); <-ent> <-ant> (/ənt/); <-ence> <-ance> (/əns/); <-ive> (/ɪv/) <-ism> (/ɪzəm/) <-age> (/ɪdʒ/); <-ture> (/tʃə/) <-ure> /ʊə/ <-our> <-or> <-er> (/ə/)
12	<-able> (/əbl/) <-ably> (/əbli/) <-ible> (/ɪbəl/) <-ibly> (/ɪblɪ/); <-tion> <-sion> <-ssion> (/ʃən/)

Appendix 2. Contents in the phonology instruction programme

Week number	Content
1	Concepts: syllables, stress and rhythm, phonemes: stops (/b/ /p/ /t/ /d/ /k/ /g/)
2	Fricatives and affricates (/f/ /v/ /θ/ /ð/ /s/ /z/ /h/ /tʃ/ /dʒ/)
3	Nasals, approximants and laterals (/m/ /n/ /ŋ/ /w/ /j/ /r/ /l/)
4	Front vowels and central vowels (/iː/ /ɪ/ /e/ /æ/ /ɜː/ /ə/)
5	Back vowels (/uː/ /ʊ/ /ɔː/ /ɒ/ /aː/ /ʌ/)
6	Diphthongs (/ɪə/ /eə/ /ʊə/ /eɪ/ /aɪ/ /ɒɪ/ /əʊ/ /aʊ/)
7	Stressed and unstressed syllables, stressed and unstressed words in a sentence
8	Strong forms and weak forms
9	Linking
10	Rhythm of English speech
11	Types of intonation and intonation units in English
12	Functions and uses of English intonation

Appendix 3. Contents of a typical lesson in the two programmes of instruction

Phonics instruction programme

- Whole-class work: students review graphemes instructed previously
- Whole-class work: teacher introduces the session's target graphemes by writing them on the board and modelling their pronunciations
- Whole-class work: students practise the target graphemes in activities such as reading words containing the target graphemes. Students are also asked to think of other words containing these graphemes. Teacher gives feedback on pronunciation and provides oral practice through choral repetition.
- Whole-class work: teacher shows some unfamiliar words containing target graphemes and asks students to segment them into graphemes. Then teacher asks students to blend the individual sounds to make the word.
- Pair-work or whole-class work: students look for words containing target graphemes in stories. Teacher corrects their pronunciation.
- Whole-class work: students read the stories in chorus. Teacher corrects pronunciation.

Phonology instruction programme

- Whole-class work: students review phonemes / pronunciation skills taught previously
- Whole-class work: teacher introduces new phonemes / pronunciation skills
- Whole-class or individual work: students practise the target phonemes / pronunciation skills in activities such as reading and listening. Teacher gives feedback on their pronunciation and provides oral practice through repetition.
- Pair-work or whole-class work: students listen to the recording of a short passage and answer some questions. Teacher provides feedback on their pronunciation.
- Whole-class or individual work: students imitate the recording of the short passage in chorus. Teacher gives feedback on pronunciation.
- Pair-work: students practice phonemes / pronunciation skills in dialogues. Teacher provides feedback on their pronunciation.

Appendix 4. Stimuli in the phonological decoding test

Form A	bab, op, dee, bim, tay, yee, pog, shum, plip, dud's, whie, bufty, vunhip, knaf, twem, adjex, yeng, laip, zirdn't, straced, cedge, wrey, whumb, knoink, bafmotbem, monglustamer, pnir, ceisminadolt
Form B	bab, op, ree, raff, dat, glack, hend, weaf, chur, tayed, ful's, rejune, weat, sess, depine, wrault, throbe, gouch, brecked, darlanker, cigbet, mancingful, squow, cyr, quiles, untroikest, pelnidlum, byrcal

Appendix 5. Stimuli in the vocabulary learning tasks

Form A	argent, doodah, ploidy, cantor, stooge, augean, burlap, tisane, sulcus, rheumy
Form B	maenad, orrery, ruddle, precis, ribose, mayhap, zeugma, turgor, callus, zephyr