

EMPIRICAL STUDY **Early Language Learning: The Impact of Teaching and Teacher Factors**Suzanne Graham , Louise Courtney, Theodoros Marinis, and Alan Tonkyn

University of Reading

This study examined the progress in lexical and grammatical knowledge among 252 learners of French in England across the last two years of primary education and into the first year of secondary school in relation to teaching and teacher factors. It compared linguistic outcomes from two different approaches, one which emphasized oracy and the other which combined literacy with attention to oracy development. We also explored the relationship between linguistic outcomes and other teaching/teacher factors: teaching time, teacher level of French proficiency, and teacher level of training in language instruction. Learners completed a sentence repetition task and a photo description task, making small but statistically significant progress in both grammatical and lexical knowledge between test points. While teaching approach had little impact on such progress, other teaching and teacher factors did, particularly the French proficiency level of the primary school teacher and the amount of teaching time devoted to French.

Keywords early language learning; teaching time; teaching approach; teacher proficiency; teacher training

This project was funded by the Nuffield Foundation. The views expressed are those of the authors and not necessarily those of the Foundation.



This article has been awarded an Open Materials badge. All materials are publicly accessible in the IRIS digital repository at <http://www.iris-database.org>. Learn more about the Open Practices badges from the Center for Open Science: <https://osf.io/tvyxz/wiki>.

The copyright line for this article was changed on 8 December 2017 after original online publication.

Correspondence concerning this article should be addressed to Suzanne Graham, Institute of Education, University of Reading, Reading, RG1 5EX, United Kingdom. E-mail: s.j.graham@reading.ac.uk

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Introduction

This study investigated the nature of the progress made by young learners of French as a foreign language in England across the last two years of primary education and the first year of secondary school. It also explored the extent to which learning outcomes are related to teaching and teacher factors within the primary setting: teacher French language proficiency, teacher level of training in language teaching, teaching time, and teaching approach, namely, the focus of instruction. We considered two approaches: “oracy,” where the teaching emphasis is predominantly on speaking and listening development, and “literacy,” where reading and writing are combined with attention to oracy.

An investigation into the progress made by young learners in classroom settings is timely and important because in recent years a growing number of countries have lowered the age at which instructed foreign language learning begins (Murphy, 2014) in the belief that an earlier start will lead to better learning outcomes. Concerns have been expressed, however, that such policy initiatives have been based on the extrapolation of findings from language learning in naturalistic settings to language learning in classroom contexts (Murphy, 2014). Such extrapolations seem to assume that both types of learning are identical and that an early start in classroom instruction will automatically lead to rapid, effortless learning. On the contrary, any age advantage reported for naturalistic foreign language learning (e.g., discussed in Muñoz, 2008) has not been found in rigorous, longitudinal studies in instructed contexts, with a later start (e.g., at age 11), in fact, often resulting in faster, more efficient learning, as was found, for example, for learners of English in Spain (Muñoz, 2006), learners of French in England (Myles & Mitchell, 2011), and learners of English in Germany (Jaekel, Schurig, Florian, & Ritter, 2017).

Furthermore, high variability in learning outcomes for young learners has been reported for instructed settings (e.g., see findings from Europe reported in Enever, 2011a), suggesting that factors other than age contribute to language learning for such children. Additionally, the question of age and its relevance in instructed contexts need to be considered with reference to the ways in which young children are believed to learn a language. DeKeyser (2003) argues that children draw on implicit mechanisms (acquiring grammatical structures through exposure alone), leading to better ultimate attainment, as “many elements of a language are hard to learn explicitly” (p. 335). Older children and adults, by contrast, learn explicitly, drawing on “native language knowledge” (p. 334); greater cognitive maturity enables them to progress more quickly. Furthermore, to be able to draw on implicit mechanisms, young learners require vast amounts of time and input. While such amounts are available in naturalistic

contexts, such as residence in the country where the language is spoken, this is far from being the case in instructed foreign language classrooms (DeKeyser, 2003; Muñoz, 2006, 2008, 2014).

If it is accepted that young learners learn implicitly and that this is a slow process requiring plentiful and high-quality input, the amount of exposure that they receive in instructed settings is of prime importance, more important than the age at which instruction begins, with the consequence that amount of exposure “never ceases to be a determinant factor” in learning outcomes (Muñoz, 2014, p. 466). This is also likely to be the case for the quality of that exposure, which may vary according to the teachers’ level of foreign language proficiency, their pedagogical skills, and the teaching methods they employ. Nevertheless, while there has been plentiful research on the effects of age on language learning, much less is known about the relationship between amount and quality of language exposure, on the one hand, and learners’ language development, on the other, in classroom settings. The present study sought to address that gap and to explore how variation in teaching and teacher factors relates to learners’ grammatical and lexical development.

The context of this study is the learning of French in England, a country that has followed the global trend toward an earlier start for language learning, albeit more recently. Foreign language instruction became a compulsory element of the National Curriculum at the primary school level in September 2014. Learners are required to make “substantial progress” in one language during the last 4 years of primary education (Department for Education, 2013, p. 1), that is, from Year 3 (ages 7–8) to Year 6 (ages 10–11). At the same time, annual surveys indicate a great deal of variability across schools in the amount of lesson time allocated to language learning and in teachers’ levels of language proficiency and training (Tinsley & Board, 2016). Variability in teaching and teacher factors is also found both across and within other countries (Enever, 2011b, commenting on Europe), and the relative absence of the target language in the environment within England and in other Anglophone contexts makes it particularly important to understand how such variability affects learning outcomes.

Variability in primary language provision is also likely to have implications for learners’ move to secondary or high school, as it means that learners in England enter secondary school with widely different levels of foreign language proficiency (Tinsley & Board, 2016). This range poses a great challenge to language teachers, which they may seek to address by simply reteaching what was meant to have been covered in earlier years (Office for Standards in Education, 2011). Such issues and practices occur not only in England but also elsewhere, for example, in the United States and Australia (Lo Bianco, 2009;

Pufahl & Rhodes, 2011); this also occur in other curriculum areas, such as mathematics and reading (Galton, Gray, & Ruddock, 1999), albeit to a lesser extent because the core nature of those areas means that variability in teaching during primary school is less pronounced. Reteaching may account for findings across a range of contexts, suggesting that learners' progress slows across the primary to secondary school transition (e.g., Hill, Davies, Oldfield, & Watson, 1998 [Australia]; Low, Brown, Johnstone, & Pirrie, 1995 [Scotland]). By contrast, and more positively, a recent study by Courtney (2014) of 26 learners of French in England across the primary–secondary transition did find evidence of progress in breadth of vocabulary, gender assignment, and verb morphology, although progress was slow and there was much individual variation. While Hill et al. (1998) and Courtney (2014) were relatively small longitudinal studies, Low et al. (1995) had a much larger sample, but used cross-sectional data, making it difficult to compare findings across the studies and possibly contributing to the apparent contradictions in their findings.

There is also evidence of some plateauing at the end of primary school, as Cable et al. (2010) found for target language phonology and listening in a study of learners of French in England across the last 4 years of primary education (Years 3–6). Slower progress in Year 6 may relate back to teacher factors in primary school, particularly to teachers' subject knowledge, which may be inadequate to deal with more than the beginner level. In turn, a lack of progress at the end of primary school may then form a shaky foundation for learners as they move into secondary school. Making progress across the primary–secondary transition is likely to have implications for subsequent motivation and success in language learning, both areas in which England has persistent problems (Tinsley & Board, 2016), along with other countries such as Australia (Lo Bianco, 2009).

The Role of Teacher and Teaching Factors

Amount of Instruction

While outcomes in early language learning have often been explored from an individual differences perspective, there is also evidence of the importance of teaching and teacher-related factors: first, teaching time (or amount of exposure learners receive), which has been an area of importance to researchers since at least the 1970s (Lightbown, 2014) and which has been investigated from a number of different angles in relation to early language learning. For example, Myles and Mitchell (2011), in a study of learners 5, 7, and 11 years old in England, established that vocabulary learning was influenced by amount of raw teacher input, that is, how often learners encountered the linguistic

items on which the project tests were based, for all ages. Frequency of lessons was reported to have influenced linguistic outcomes (reading, writing, and listening) in a large study ($N = 20,804$) of Year 6 learners of English in Hungary by Nikolov (2009). Although Nikolov claimed that the number of weekly lessons learners experienced was positively correlated with scores on the language tasks used, the absence of detailed results in the reporting of this study make the correlation difficult to interpret. Furthermore, it is unclear whether the relationship between lesson frequency and outcomes was stronger for the Year 6 learners than for the Year 10 learners also targeted in the study.

Looking at total amount of lesson time and length of lessons, Genelot (1997), in a study of over 1,000 young French learners of English, reported a positive relationship between total amount of lesson time and learning outcomes, as well as a more positive effect of shorter, more frequent lessons compared with longer, less frequent ones. By contrast, there is also evidence that intensive periods of instruction may be more beneficial than a drip-feed (little and often) approach. Intensive instruction may be needed at the beginning of language study to allow learners to reach the basic level of communicative competence on which further development can be built. This is the conclusion reached by Netten and Germain (2008) in Canada, comparing core French (daily lessons of 30–50 minutes) with an intensive approach (a 5-month period of intensive exposure to French across 65–70% of the school day), followed by a return to a more regular curriculum (typically two 80-minute sessions per week). Intensive French aims to develop spontaneous oral and written communication, with literacy skills developed alongside oracy, for which, the authors argue, the longer lesson time is required. Evaluating outcomes from both core and intensive French, Netten and Germain reported that learners who experienced the intensive approach developed spontaneous communication skills, but core French learners did not. While it must be acknowledged that the intensive French program discussed by Netten and Germain differs from the type of instruction offered in the context of the present study, there are still, arguably, implications that can be drawn from it with respect to the importance of time for learning.

Teaching Approach: Literacy and Oracy

Linking oracy and literacy skills, according to Netten and Germain (2008), is central to the success of intensive French. The introduction of reading and writing into early language learning is not universally supported, however. Indeed, in England, an “oracy first” approach to second language (L2) instruction is widely used (Cable et al., 2010), possibly based on the premise that instructed L2 learning should follow what happens in first language (L1) development—that

is, language is first acquired orally and only later are grapheme-phoneme correspondences taught. Oral and aural instruction involves more implicit learning and, hence, might be deemed to be more appropriate for younger learners.

Although literacy-based activities may be considered more suitable for older learners, it is possible that they can also benefit younger learners. Evidence from immersion studies suggests that oral input alone may lead to fluency but inaccurate production, as Harley and Swain (1984) found for past-tense formation among Grade 1 English-speaking learners of French in Canada. They argued that learners are less likely to notice or attend to grammatical forms from oral input, especially when grammatical features have lower levels of phonological salience. The authors also commented that “written input of some kind would be helpful in drawing attention to phonologically non-salient segments” (p. 295). This may particularly be the case for a language such as French, in which markers of gender and adjectival agreement are not very salient in oral input (Courtney, 2014).

Vocabulary development also potentially benefits from written input rather than oral input alone, according to a range of evidence reviewed by Hu (2008), who argues that a focus on orthographic forms helps to “‘fossilize’ the L2 speech signals” (p. 823) in the input, that is, it makes them more accessible and therefore more likely to be processed effectively. Beginning learners of a L2, especially those with weaker phonological awareness and, hence, poor speech perception, may find it hard to learn from oral input alone because they have difficulty in “constructing accurate, detailed phonological representations in the process of abstracting a stable specification of the sound structure of the new word from the input” (p. 825). Thus, presenting the written form alongside the oral form could result in a clearer and more durable memory representation, although a counterargument might be made that presenting language in more than one modality presents learners with a heavier cognitive load.¹ In Hu’s study of Grade 3 Chinese-speaking children in Taiwan, learners were taught novel English words, both with and without written forms. Half the sample (37) had higher levels of L1 phonological awareness while the other half had lower levels. Learning was found to be better in the written condition, but learners with higher phonological awareness benefited the most from the written input. Hu hypothesized that the benefit for learners with low phonological awareness could be enhanced with more explicit instruction in L2 phonics, highlighting the importance of the quality over the type of teaching.

Written input can perhaps also provide more repeated encounters with lexis and grammatical structures, also potentially covering a wider range than is possible from the teacher’s oral input alone (Lightbown, 2014; Porter, 2014).

This input compensates, perhaps, for any shortcomings in the range and accuracy of the teacher's language proficiency, which may then have benefits not only for learners' literacy development in the L2, but also for their oral development. Indeed, two studies indicate that oracy and literacy can develop side by side. The first, by Drew (2009), investigated the impact of an Early Years Literacy Programme on the development of learners of English (age 8) in Norway. The program included extensive reading of illustrated, graded books that featured systematically high-frequency vocabulary. An experimental group of 57 learners received lessons combining periods of extensive and differentiated reading with oral, communication-based activities. Learners also read at least one book a week at home with the help of parents. Their performance on pre- and posttests in listening, speaking, reading, and writing was compared with that of 58 learners experiencing lessons through largely whole-class teaching, choral repetition, reading dialogues aloud in groups, and simple writing. The experimental group made more progress than the control group on all skills, but especially in the listening and oral tests. More recently, and in the same context as the present study, Porter (2014) conducted an action research project with primary school-age learners of French in England and found that literacy work, including both reading and writing, allowed learners across the attainment range to make progress orally as well as in literacy. Although both studies indicate that combining literacy with oracy can be beneficial across skills, Drew's findings suggest that the experimental group may also have received greater amounts of input than the control group from the out-of-class work, while the absence of a control group in Porter's study makes it more difficult to assess the impact of literacy work.

Teacher Training

Quality of input is likely to be of as much importance as quantity and type of input, particularly with regard to the language proficiency and pedagogical skills of teachers. Neither input quality nor input quantity has received extensive research attention, and studies have also tended to be small and cross-sectional. For example, the impact of teacher training was considered in a study by Mihaljević Djigunović (2009), who looked at the relationship between attitudes and learning conditions for children learning English in Croatia. Learning conditions included teaching time, teacher training, and class size, the latter being a factor which, in a review of studies of early language learning in Europe, Blondin et al. (1998) found to have negligible impact on outcomes. Mihaljević Djigunović reported that "less favorable" conditions negatively impacted learners' attitudes toward and sense of competence in learning English,

although few details were provided regarding the exact nature of the conditions in which learners were taught. For example, in the four less-favorable classes, “the teacher of English *might* not have had much training” (p. 78, emphasis added). Similarly, Szpotowicz (2009) investigated vocabulary acquisition over 3 to 5 days with 67 learners of English in Poland. Although the main focus was on the learnability of different types of vocabulary, a teacher effect was also identified, with learners taught by the least qualified of four teachers achieving the lowest scores on tests of recall and recognition. This was in spite of all four teachers using the same vocabulary teaching methods prescribed by the researcher.

Teachers’ Language Proficiency

It is unclear, however, whether teachers in Szpotowicz (2009) also varied in English language proficiency. Teacher language competence is likely to have an impact on learning outcomes, with Muñoz (2006) stating that not only does exposure need to be intense but that it also needs to “provide an adequate model” (p. 34). Teachers’ linguistic skills are likely to be especially relevant to the acquisition of grammatical features, particularly within usage-based theories of language acquisition, which emphasize the importance of the linguistic environment (see Murphy, 2014, for an overview) for grammatical development. Frequency and consistency (in relation to form–function mappings) of grammatical features in the input are likely to influence how well such features are acquired (Murphy, 2014). Arguably, teachers with lower levels of proficiency in the target language are less able to provide such frequency and consistency. The optimal level of teacher language proficiency is, however, more difficult to determine. Native speakers, particularly if they lack training in language teaching (Walkinshaw & Duong, 2012), may be less able to simplify their output to make it comprehensible to learners (Krashen, 1982).

Overall, studies that have investigated the impact of teachers’ level of language proficiency are limited in number, perhaps, as Unsworth, Persson, Prins, and De Bot (2015) argue, it is a rather sensitive topic. Some studies relate to bilingual rather than foreign language settings; certain tentative conclusions from them can, however, be drawn. For example, Bowers and Vasilyeva (2011) reported that for early learners of English in a bilingual setting, the total number of words spoken by their native-speaker teachers during a period of 90-minute observation (audio recorded) was positively related to learner’s receptive vocabulary knowledge, but that the average number of words per teacher utterance was negatively correlated with learning outcomes. The authors interpreted this as an indication that beginning language learners need

exposure to relatively uncomplicated speech and can only take advantage of native-speaker input once they are beyond the very beginner stage. Perhaps these findings also suggest that teachers need to know how to make their speech comprehensible and suitable for the learners they are teaching. Native-speaker input may, however, sow the seeds for later development, according to Aukrust (2007), who found that in another bilingual setting, amount, diversity, and discourse complexity of teachers' speech only predicted preschool learners' language outcomes 2 years later.

Amount and quality of input may, furthermore, interact with each other. This issue has been the subject of investigation in two studies conducted in foreign language instructed settings in the Netherlands. The most recent, by Unsworth et al. (2015), explored the impact of weekly lesson time for language learning and teacher language proficiency on the development of vocabulary and grammar skills of Dutch children learning English in their first and second year of study. Learners constituted two groups: 168 learners in early English schools (receiving up to 220 minutes a week of English) and 26 age-matched children having more regular English exposure, that is, approximately 45 minutes a week. The authors found that both teaching time and teacher oral language proficiency were important factors in learners' scores for grammar and vocabulary development. This importance was especially the case when children received under 60 minutes a week of English instruction, and when the teachers' proficiency level was below level B on the Common European Framework of Reference (CEFR; Council of Europe, 2001). Furthermore, a regression analysis showed that teachers' language proficiency was the best predictor of outcomes for both vocabulary and grammar, a finding that the authors interpreted as evidence of the importance of lexical diversity and grammatical complexity in input gained from the teacher. Finally, teachers' language proficiency predicted grammar scores only at the second posttest, suggesting that its effects take time to emerge, a conclusion that underlines the importance of longitudinal investigations of the impact of teacher variables.

Unsworth et al. (2015) noted that their study did not take into account other issues that may have impacted learning outcomes, such as teaching qualifications and the kind of instruction given, which underscores the need to take account of the full range of teaching/teacher factors that may influence learning. An earlier study in the Netherlands by Edelenbos and Suhre (1994) is one of the few to explore the teacher variables of teaching experience and spoken fluency, language teaching qualifications, and teaching time, alongside teaching approach (either broadly communicative or grammar-focused courses, as ascertained through content analysis). Data on reading, writing, listening,

vocabulary, grammar, and spoken fluency were collected from 2,116 pupils in 112 schools. Amount of lesson time was significantly and positively correlated with all scores except learners' spoken fluency. Possessing a teaching qualification was the most important teacher factor for predicting outcomes (in vocabulary, grammar, spoken fluency, and listening), with teacher spoken fluency positively related only to learners' grammar scores. Once teacher and socioeconomic status variables were controlled for, however, the only difference attributable to the course followed was on the grammar test, where, perhaps unsurprisingly, learners following a course with a heavier emphasis on explicit grammar had the higher scores. Additionally, as Driscoll, Martin, Graham-Matheson, Dismore, and Sykes (2004) point out, the teachers in Edelenbos and Suhre's study who used the more grammar-focused courses were also the most qualified and experienced teachers and had more English language teaching time available to them. A further limitation of the study, however, is that it was cross-sectional and thus could not give full insight into how teaching and teacher factors impact learners' progress over time.

The Current Study

Variability in teaching and teacher factors, such as those noted above, is pertinent to the context of England because, compared with what children in other European countries experience (Enever, 2011b), learners in England receive limited amounts of language instruction. In an annual survey of language teaching provision of approximately 600 primary schools (Tinsley & Board, 2016), 48% reported allocating 30–45 minutes per week to language teaching, 32% between 45 minutes and 1 hour, and just 15% between 1 and 2 hours. Additionally, about one third of schools reported that the staff members teaching a foreign language had only General Certificate of Secondary Education (GCSE) level of language competence (i.e., CEFR level A) or below. Teachers with native- or university degree-level competence (i.e., CEFR level C2) were reported by 16% and 29% of schools, respectively. Nearly 60% of responding schools reported that lack of staff expertise in language pedagogy was a challenge for them. These less-than-ideal conditions for language learning are not unique to England. A survey of 142 countries by Copland, Garton, and Burns (2014) suggested that concerns about teaching and teacher factors are widespread. Nevertheless, the fact that in England the classroom is the only real contact that most learners have with the foreign language they are expected to learn, which in most schools is French (Tinsley & Board, 2016), means that teaching time, teacher language proficiency, and teacher language pedagogy training are likely to be of particular relevance.

Regarding teaching approach or focus of teaching, curriculum documents in England relating to primary languages make strong statements about the importance of including literacy in early language instruction. Thus, the Key Stage 2 Framework for Languages (Department for Education, 2005), a non-statutory but still widely used curriculum document (Cable et al., 2010; Porter, 2014), includes both an oracy and literacy strand and presents them as being of equal importance. Nevertheless, Cable et al. found that oracy received more attention than literacy in many classrooms in England.

In summary, previous research has indicated that amount of teaching time, teacher language proficiency, teacher training, and teaching approach (i.e., focusing predominantly on oracy or attending to literacy as well) are likely to have an impact on the language development of young learners, but in potentially complex ways. It is not clear, for example, which factors are the most important, at which stage of learning, and whether they interact with one another. Overall, given that large proportions of young learners in England are being taught in conditions that are at odds with what research suggests as being optimum for language acquisition, it is important to gain greater understanding of how learning outcomes vary with teaching and teacher factors, and from a longitudinal perspective. Furthermore, the somewhat contradictory findings regarding the nature of learners' progress across the primary–secondary transition call for further exploration of the nature of that progress.

Therefore, the current study aimed to address these issues by using a longitudinal design to investigate the following research questions:

1. To what extent do children make progress in their knowledge of French (measured through tests of vocabulary and grammar) across Years 5, 6, and 7?
2. To what extent is children's knowledge of French across Years 5, 6, and 7 related to (a) teaching and teacher factors in primary school (teaching time, teacher French proficiency, teacher level of training) and (b) teaching approach (oracy vs. literacy)?

Our measures of grammatical development included three grammatical features of French that are part of the primary languages curriculum in England (Department for Education, 2005, 2013): article–noun agreement, adjective–noun agreement, and simple present-tense verbs (considered together to provide a broad measure of grammatical accuracy). Curriculum expectations are not that teachers should engage in formal, explicit grammar teaching but rather that they should develop learners' "knowledge about language" (Department for Education, 2005) by drawing their attention to features such as gender class

Table 1 Data collection points and number of participants (total sample and subsample by teaching approach)

Test point	Total sample			By teaching approach		
	Male	Female	Total	Oracy	Literacy	Total
1. Year 5 (summer)	119	133	252	73	103	176
2. Year 6 (spring/summer)	114	127	241	68	100	168
3. Year 7 (autumn)	75	89	164	54	77	131

as part of more communicative, game-based activities—practices observed by Cable et al. (2010) to be common in primary classrooms in England. It might be claimed that grammatical gender is difficult for English-speaking learners of French to acquire.² There is, however, evidence that accuracy in gender assignment increases at a statistically significant level between school Years 5 and 7, as Courtney (2014) found, reaching a 69% accuracy level. We therefore considered it appropriate to include it within our broader assessment of grammatical development in relation to teaching and teacher factors.

Method

Research Design

The study was a “natural experiment” (Shadish, Cook, & Campbell, 2002, p. 12), in which the outcomes of naturally occurring phenomena were compared. It tracked learners’ linguistic development in French from the penultimate year of primary school, Year 5 (summer term, Test Point 1), to Year 6 (early summer term, Test Point 2), and into Year 7, the first year of secondary school (halfway through the autumn term, Test Point 3). At all test points, learners completed a sentence repetition task and a photo description task, both described below. Data collection stages are shown in Table 1.

To address the second research question, we looked at the two following teaching approaches:

1. a predominantly oral approach, which focuses on developing speaking and listening skills with very little emphasis on literacy (oracy) and
2. a more literacy-based approach, where reading and writing activities are integrated into instruction, alongside oracy skills (literacy).

Although we were contrasting two approaches that might be termed “oracy” and “oracy + literacy,” for brevity and clarity we refer to the approaches as oracy and literacy, respectively. Furthermore, we use the term *approach* to signify the

relative emphasis placed by teachers on the development of oracy and literacy skills and the relative amount of time spent on activities involving oral/aural or written language.

Learners

Data were initially collected from 254 learners across nine primary schools. The data from two learners were subsequently excluded from the analysis, as their English literacy scores were greater than three standard deviations below the mean, giving a Year 5 sample of 252. Numbers of learners involved in the study varied across test points, as indicated in Table 1. Attrition at the start of Year 7 arose for two main reasons: Some learners transferred to a secondary school where the language taught was Spanish or German rather than French; and several secondary schools required us to test learners outside of lesson time, and we hence had to rely more on learners' willingness to attend afterschool testing sessions (not all learners attended such sessions). Those withdrawing in Year 7 had test scores at Test Points 1 and 2 that did not differ significantly from those of learners who remained in the study at Test Point 3 (as ascertained through independent-samples *t* tests, with *p* values ranging from .16 to .99), giving us confidence that the remaining students were not atypical of the sample as a whole.

All learners spoke English at home (as ascertained through a parent questionnaire). One learner also spoke a language in addition to English at home (Nepali), but had levels of English literacy within three standard deviations of the sample and the national average. Nepali does not mark gender or adjective agreement as French does, and so the learner was not felt to be at an advantage for the targeted measures of French. This was further supported by the fact that this learner's scores on all measures of French were within one standard deviation of the sample mean. For these reasons, it was decided to retain this learner's data for analysis. Only 18% of children were reported to have had contact with French outside of school (in all cases very slight). Information about learners' English literacy levels was obtained from their primary school for reading comprehension and writing (including imaginative and factual writing, spelling, and grammar).

In England at the time of the study, learners' English literacy attainment was expressed through nationally applied National Curriculum levels and sublevels, or descriptors (e.g., 1a, 1b, 1c) based on teacher assessments and tests, with learners in our study scoring between level 1c (lowest) and 5a (highest). For the study, sublevels were converted to a point scale from 1 (1a) to 13 (5a). The scores thus obtained for reading and writing were added together and the

mean calculated, giving each learner a combined English literacy score out of a possible 13. Learners were all in their third year of learning French at the start of the project and in the penultimate year of primary education (Year 5, age between 9 years 10 months and 10 years 10 months). They were all in classes of fewer than 30 learners, with most classes comprising between 25 and 30 learners.

Literacy and Oracy Divisions of Schools

Prior to the start of the project, we distributed a questionnaire to 35 schools in the south of England, which asked teachers of primary French about the kind of activities they personally used with Year 5 learners and the frequency with which they used them. From the replies received, nine included follow-up contact details from teachers who were interested in being part of the main study. Details of these nine schools and the one teacher per school involved in the study are given in Appendix S1 in the Supporting Information online. The Key Stage 2 Framework for Languages, which is a nonstatutory document that sets objectives and related teaching activities in the areas of oracy and literacy across Years 3 to 6 (Department for Education, 2005), was used to create the questionnaire items (15 literacy, 14 oracy). Sample questionnaire items included “Pupils listen to a story in the language but don’t see the words at the same time” (oracy) and “I read stories to learners and they follow the words on the board or from a big book” (literacy).

Additionally, one question asked whether teachers drew any attention to grammatical features, such as gender of words. All teachers reported doing so. In order to gain a broad, overall picture of the teaching approach adopted, we asked teachers to indicate how often they used each activity with learners in Year 5 French classes, and space was provided at the end of the questionnaire for them to list any other activities not mentioned. Finally, teachers were asked to indicate what proportion of lesson time was devoted to listening, speaking, reading, and writing, respectively.

The percentage of literacy and oracy activities that teachers reported using in all or most Year 5 lessons was calculated. A literacy score (1 = low, 5 = high) was then allocated based on the following percentages of questionnaire literacy activities used in all or most lessons (1 = 0–20%, 2 = 21–30%, 3 = 31–40%, 4 = 41–50%, 5 = 50% or more). After gaining this broad view of the participating teachers’ instructional focus in Year 5, we wanted to be sure that the approach learners experienced did not change when they moved into Year 6. Therefore, once the study was underway, participating teachers (one per school) were observed teaching the project learners in Year 6, once

Table 2 Literacy scale

Scale value	Description
1	Year 3 reading activities observed but no writing activities; for example, recognition of familiar words in written form, making sound-spelling links, reading aloud simple words.
2	Year 4 reading activities observed but no writing activities; for example, reading and understanding familiar phrases, following a short written text that is read aloud, reading aloud phrases.
3	Year 3 and 4 reading activities in addition to Year 3 writing activities observed: as in 1 and 2, plus writing single words.
4	Year 5 and 6 reading activities in addition to Year 3 and 4 writing activities observed; for example, reading short texts individually, writing words and phrases.
5	Year 5 and 6 reading in addition to Year 5 and 6 writing activities observed; for example, reading short texts, including authentic texts, writing sentences on a range of topics using a model.

per school, using a lesson observation schedule, which also allowed for more in-depth examination of classroom activities than had been gained through the questionnaire. This observation confirmed that for each school, the teaching approach remained constant across the 2 years and gave us a well-rounded picture of the focus of instruction each group of children received across both Year 5 and Year 6. We also requested Schemes of Work (curriculum plans) from schools in order to corroborate information gained from the questionnaire and observation and to place the observation within a broader context. The types of literacy activities observed were recorded and analyzed with reference to the Key Stage 2 Framework objectives for literacy (Department for Education, 2005). We used this framework to create a literacy scale from 1 to 5, as illustrated in Table 2.

Thus, a school receiving a score of 1 would only be including the type of reading activities aimed at Year 3 in the teaching of Year 6 learners. The final observation score was then added to the questionnaire score to give each school a total literacy score out of 10. Schools with a borderline oracy/literacy score (i.e., scoring at the mean and median points of 6 and 6.5) were excluded from the analysis of the impact of teaching approaches on learning outcomes but were included for other analyses. This gave a subsample consisting of three schools in the oracy group (Schools 2, 5, and 7) and three in the literacy group (Schools 1, 3, and 9); the numbers of participants are shown in Table 1.

Schools were matched on key indicators, including percentage of pupils claiming free school meals (FSM), percentage of pupils with special educational needs, and percentage of pupils with English as an additional language (EAL). All schools were in the Department for Education's low percentage band for EAL and FSM (the latter a measure of deprivation widely used in the United Kingdom). Information on teaching and teacher factors (weekly amount of teaching time, teacher level of French, teacher level of training in teaching languages) was gained through a short questionnaire. Most teachers³ were relatively highly qualified linguistically and pedagogically, compared with the national averages (Tinsley & Board, 2016), discussed previously. As shown in Appendix S1, approximately two-thirds of teachers had native or degree-level competence in French and had had postgraduate training in language teaching. The teaching time for French ranged from 15 minutes a week for one school to 60 minutes in two schools, with the remainder at around the national average of 30–45 minutes per week (Tinsley & Board, 2016). When learners were in Year 7 (secondary school), they all received similar amounts of French teaching per week (approximately 2 hours).

Materials

We designed and piloted two tests to assess knowledge of grammatical features and vocabulary: a sentence repetition (SR) task and a photo description (PD) task, with the same tests used at each test point.

SR Task

SR tasks (also known as elicited imitation) are frequently used to assess the language development of young learners (Marinis & Armon-Lotem, 2015) because they are believed to tap into their implicit knowledge (Erlam, 2006) and have been found to be good indicators of grammatical development (e.g., Devescovi & Caselli, 2007). It is claimed that learners can only repeat structures that they have acquired and that SR tasks thus give a window into the underlying linguistic competence of the learner (e.g., Wu & Ortega, 2013). As such, SR tasks are believed to be “reconstructive” as opposed to purely imitative in nature (Jessop, Suzuki, & Tomita, 2007, p. 215). A number of recent reviews conclude that SR tasks discriminate reliably between learners of different levels in different knowledge areas. A recent meta-analysis of 76 elicited imitation studies by Yan, Maeda, Lv, and Ginther (2016) targeting investigations across syntax, morphosyntax, lexis, and phonology found strong correlations between repetition scores and other measures, leading them to conclude that SR tasks offer sensitive, valid measures. Other advantages of SR tasks are that they

permit the examination and tracking of a range of very specific linguistic items over time, which corresponded well to the aims of our study.

In SR tasks, learners hear sentences containing the target items, which they then have to repeat verbatim. Sentences need to be long and complex enough for learners to have to analyze rather than just parrot what they hear, but not so long as to place a heavy burden on memory. The issues of length and complexity have been widely discussed in the literature (e.g., Devescovi & Caselli, 2007) and are relative to the age, memory capacity, and proficiency level of the participants, with no consensus as to what is ideal (Yan et al., 2016). Choosing the right length and complexity for learners is often an empirical question and requires piloting to ensure that learners do not show a ceiling effect that could indicate parroting, or a floor effect that could suggest the length and complexity exceeds the participants' processing capacity.

The SR task used in this study (based on Marinis & Armon-Lotem, 2015) consisted of 18 sentences, with 6 each for the following grammatical areas: article–noun agreement, adjective–noun agreement, and simple present tense. These grammatical areas were targeted through 19 lexical items of one or three syllables in length, through which we also sought to assess learners' lexical knowledge (see Appendix S2 in the Supporting Information online). Sentences had 7–10 syllables, a range chosen based primarily on a review of the literature (e.g., Campfield & Murphy, 2014; Erlam, 2006) and then piloting with a small number of children, which showed that some children had a ceiling effect in sentences that were shorter than 7 syllables and a floor effect in sentences that were longer than 10 syllables. Learners saw a picture depicting the target item at the same time as hearing the phrase in order to focus them on meaning, which maximizes the possibility of reconstructive production rather than rote repetition (Erlam, 2006). The reconstructive nature of the task is also maximized and dependency on memory reduced by having a delay of about 3 seconds between hearing the sentence and repeating it (Erlam, 2006); however, for learners of very low proficiency, this delay places a heavy load on working memory. Piloting suggested that inserting an artificial delay meant learners could not do the task at all. Therefore, in the main task, no delay was inserted. Sentences were recorded with clear articulation by the second author (university degree–level proficiency in French).

The lexical items targeting the three grammar areas were placed initially, medially, and finally in the sentences, to control for order and recency effects (following Jessop et al., 2007). To minimize possible order effects, two sequences of the task were created, and learners were randomly assigned to complete one of these sequences. Lexical items were selected by consulting

commonly used primary French teaching resources and primary French practitioners regarding what children in Year 5 would be likely to have covered in their French lessons. The SR task was scored out of 56 (see below for details), with 28 points available for grammar and 28 for vocabulary targets (see Appendix S2 in the Supporting Information online).

PD Task

The PD task aimed to gain an additional perspective on learners' grammatical and vocabulary knowledge (focusing on the same target areas) and thus to help corroborate the insights gained from the SR task, as recommended in the elicited imitation literature (e.g., Jessop et al., 2007). It was in two parts: (a) focusing on article–noun–adjective agreements and (b) focusing on present-tense verb use. For the first part, pictures of colored objects were presented, and participants were asked to describe them, prompted by the question, *Qu'est-ce que c'est?* (“What is it?”). For the second part, pictures depicting actions were shown, and participants were asked to say what was happening in the picture, prompted by the question *Qu'est-ce qu'il fait?* (“What is he doing?”). In the PD task, 17 lexical items were targeted, overlapping with those in the SR task (divided between the three grammatical areas). The PD task was scored out of 54, with 30 points for grammar and 24 for vocabulary (for details of task content and scoring, see Appendix S3 in the Supporting Information online).

Procedure

Learners were tested individually during French lesson time in Years 5 and 6 and, in most cases, after school in Year 7, using a laptop in a quiet room. Pictures for each item in the SR task were shown on the laptop by a researcher, and the learner listened to each sentence once through headphones before giving a response (repeating the sentence), which was recorded using *Audacity* and an external microphone. Similar procedures were followed for the PD task, except that learners saw pictures and were asked to say in French what they saw. The researcher followed a scripted protocol, and response to the learner was limited to a simple phrase of encouragement after each attempt in the tasks. All appropriate information and consent procedures were followed. As a token of our appreciation for taking part in the study, learners in Year 7 were given a low-value gift.

Information regarding teaching and teacher variables was drawn from the teacher questionnaire, in which primary school teachers had given information about the number of minutes per week spent on French in Years 5 and 6,

their own French proficiency (i.e., their highest level of French qualification), and their training in language teaching (see Appendix S1). For French qualifications, teachers chose one from the following options: no formal French qualification, GCSE (CEFR level A), A level (CEFR level B2), university degree (CEFR level C2), and native speaker. Participating teachers' proficiency covered all these levels except A level. For teaching qualifications, questionnaire options were: no training, some training within a generalist teacher training course, and postgraduate qualification specializing in foreign languages. All these levels were represented among the primary school teachers in the study. Teacher training and language proficiency variables remained constant across Years 5 and 6 in each primary school, as was the case for teaching approach.

Data Analysis

All sentences were fully transcribed, and mispronunciations were transcribed phonetically using English spelling conventions, for example, *la apa* (produced instead of *le lapin*, "the rabbit," in the SR task). Indeterminate forms of the indefinite article were transcribed as (*if*). The sound quality was excellent and, therefore, inaudible words occurred very rarely and were marked with *xx* for one word or *xxx* for multiple words. We scored only target items, rather than using a more holistic scoring, because our aim was to gain insights into grammatical and lexical development by tracking the same grammatical forms and lexical items over time. Scoring is outlined in detail in Appendices S2 and S3; briefly, a vocabulary mark was first awarded if the required noun or verb was produced, then grammar marks were given for correct gender assignment, adjectival agreement, and simple present tense, as applicable.

Nativelike pronunciation was not required. Scores for grammar (combining all three targets) and vocabulary were calculated for each learner per task, plus a global score (grammar plus vocabulary). Scoring was carried out first by the second author; the first author (also with university degree-level competence in French) then scored 10% of learners' SR and PD tasks. An interrater reliability rate of 98% was achieved, with differences in scores resolved through discussion. The calculation of Cronbach's alpha indicated strong reliability both for the SR task ($a = .81-.84$ for vocabulary, $a = .80-.83$ for grammar, and $a = .93-.94$ for global scores) and for the PD task ($a = .84-.86$ for vocabulary, $a = .90-.91$ for grammar, and $a = .93-.94$ for global scores).

Normality and homogeneity of variance assumptions were assessed by examining histograms and normality tests (Shapiro-Wilks), which indicated that the majority of global scores from the total sample and the oracy/literacy

schools subsamples were normally distributed at each test point, except for the Year 7 SR test (total sample: $SW = .97$, $df = 160$, $p < .001$; literacy subsample: $SW = .95$, $df = 75$, $p = .006$) and Year 5 PD test (total sample: $SW = .97$, $df = 160$, $p = .001$; literacy subsample: $SW = .97$, $df = 75$, $p = .034$; oracy subsample, $SW = .94$, $df = 53$, $p = .014$). Histograms suggested, however, that deviations from normality were not severe and, following Field (2013), it was decided that parametric tests were robust enough to cope with the slight deviations from normality for these scores (with assumptions for each individual test checked and reported separately).

Separate grammar and vocabulary scores for both tests showed a greater level of nonnormality and, therefore, nonparametric statistics were applied in any analyses of those scores. Unless indicated, the alpha level was set at .05. Results for post hoc tests are reported using SPSS Bonferroni-adjusted p values. Where there were any instances of violations of sphericity (Mauchly's test), degrees of freedom were corrected using Huynh-Feldt estimates of sphericity.

Results

Linguistic Outcomes

Our first research question asked to what extent children make progress in their knowledge of French (vocabulary and grammar) across Years 5, 6, and 7. In the total sample, learners' scores increased steadily across the three testing rounds (Table 3), for global scores and separate vocabulary and grammar scores. For global scores, a one-factor repeated-measures analysis of variance (ANOVA) indicated that there was a significant effect for time for the SR task, $F(2, 318) = 118.28$, $p < .001$, $\eta_p^2 = .43$. Bonferroni tests indicated that learners made significant progress between all three test points ($p < .001$), with small effect sizes⁴ for T1–T2 ($d = 0.34$) and T2–T3 ($d = 0.44$) and a medium effect size for T1–T3 ($d = 0.79$). Similar results were obtained for the PD task, $F(1.72, 272.77) = 96.53$, $p < .001$, $\eta_p^2 = .38$, with significant progress across all three test points ($p < .001$) and small to medium effect sizes for T1–T2 ($d = 0.27$), T2–T3 ($d = 0.45$), and T1–T3 ($d = 0.70$).

Examining grammar and vocabulary separately, scores were low at all test points, especially for grammar. Nevertheless, nonparametric analyses (a Friedman test with post hoc Wilcoxon matched-pairs signed ranks tests) showed that learners made statistically significant progress over time in both grammar and vocabulary. While effect sizes for vocabulary growth were in the medium to large range, they were generally small to medium for grammar (Table 4).

Table 3 Means (standard deviations) for vocabulary, grammar, and global scores for the entire sample in Years 5–7

Test point	Vocabulary		Grammar		Global	
	<i>M (SD)</i>	<i>Range</i>	<i>M (SD)</i>	<i>Range</i>	<i>M (SD)</i>	<i>Range</i>
Sentence repetition (SR) task						
Year 5 (<i>N</i> = 252)	14.40 (4.86)	1–26	9.08 (5.12)	0–24	23.48 (9.58)	1–50
Year 6 (<i>N</i> = 241)	17.02 (5.34)	4–27	9.76 (4.72)	1–23	26.78 (9.71)	6–49
Year 7 (<i>N</i> = 164)	19.52 (4.72)	6–28	11.37 (4.59)	1–23	30.88 (8.89)	8–51
Photo description (PD) task						
Year 5 (<i>N</i> = 252)	9.52 (4.81)	0–23	4.39 (3.65)	0–18	13.91 (8.14)	0–39
Year 6 (<i>N</i> = 241)	10.98 (4.65)	0–21	5.11 (3.55)	0–19	16.09 (7.78)	0–40
Year 7 (<i>N</i> = 164)	12.52 (4.95)	0–23	7.14 (5.08)	0–24	19.66 (8.48)	4–45

Notes. In the SR task, the maximum possible scores for vocabulary and grammar are 28 each. In the PD task, the maximum possible scores for vocabulary and grammar are 24 and 30, respectively.

Table 4 Comparisons across test points for vocabulary and grammar scores (Wilcoxon matched-pairs signed ranks test, $\alpha = .008$)

Comparison	Vocabulary			Grammar		
	<i>z</i>	<i>p</i>	<i>r</i>	<i>z</i>	<i>p</i>	<i>r</i>
Sentence repetition task						
Years 5 vs. 6	9.75	.001	.44	3.12	.001	.14
Years 6 vs. 7	7.83	.001	.39	4.55	.001	.23
Years 5 vs. 7	10.22	.001	.50	6.45	.001	.32
Photo description task						
Years 5 vs. 6	8.21	.001	.37	4.50	.001	.20
Years 6 vs. 7	5.62	.001	.28	5.59	.001	.28
Years 5 vs. 7	8.28	.001	.41	6.60	.001	.32

Linguistic Outcomes and Teaching Factors

Our second research question asked to what extent children's knowledge of French across Years 5, 6, and 7 is related to (a) teaching and teacher factors in primary school (teaching time, teacher French proficiency, teacher level of training) and (b) teaching approach. Looking first at the relationship between linguistic outcomes and primary school teaching/teacher factors, a series of Spearman correlations was conducted across the total sample for grammar and vocabulary as well as global scores. As shown in Table 5, the level of

Table 5 Spearman correlations between linguistic scores and teaching/teacher variables

Score	Year	Teacher proficiency	Teacher training	Teaching time
Sentence repetition task				
Global scores	5	.126*	.098	.239**
	6	.182**	.202**	.362**
	7	.230**	.222**	.231**
Vocabulary	5	.045	.053	.158*
	6	.179**	.181**	.389**
	7	.241**	.210**	.247**
Grammar	5	.182**	.132*	.279**
	6	.175**	.216**	.303**
	7	.192*	.211**	.211**
Photo description task				
Global scores	5	.198**	.161*	.424**
	6	.165*	.140*	.392**
	7	.239**	.183*	.235**
Vocabulary	5	.180**	.147*	.388**
	6	.130*	.083	.373**
	7	-.006	.059	.121
Grammar	5	.196**	.152*	.436**
	6	.203**	.215**	.377**
	7	.420**	.274**	.288**

Note. * $p < .05$, ** $p < .01$, two-tailed.

the primary school teachers' French language proficiency and their level of training were significantly related to learners' linguistic outcomes at all points for grammar and at most test points for vocabulary, although at low to moderate levels, in part attributable to the fairly narrow range of the teachers' French proficiency and training ratings. Teaching time was more strongly correlated with outcomes than teacher variables, especially in Years 5 and 6, with Year 5 outcomes only weakly related to teacher factors. As learners moved into secondary school, however, the relationship between their scores and the level of French proficiency and training of the teacher who had taught them at primary school became stronger, particularly where grammar was concerned.

Teacher (French) Proficiency and Teacher Training

To explore the impact of primary school teacher variables further, three groups were created for teacher proficiency (i.e., French language proficiency) in order to yield groups of comparable size: GCSE or lower, degree level, or native

speaker. For teacher training, two groups were created: nonspecialist (no or limited training, within the context of a general teaching qualification) and specialist (teachers had specialized in foreign languages for their teaching qualification).⁵ Repeated-measures ANOVAs were first performed on global scores with teacher proficiency or teacher training as a between-subjects factor, test point as a within-subjects factor, and teaching time at primary school as a covariate. The mean global scores across teacher groups for the SR and PD tasks, adjusted for teaching time, are depicted visually in Appendix S4 in the Supporting Information online.

Teacher Proficiency and Global Scores

Homogeneity of variances ($p > .05$) and covariances ($p > .05$), as assessed by Levene's and Box's M tests, was established for both the SR and PD task. There was a significant main effect of test point for the SR task, $F(2, 312) = 20.21$, $p < .001$, $\eta_p^2 = .12$, and the PD task, $F(1.74, 272.09) = 28.67$, $p < .001$, $\eta_p^2 = .16$, but no significant main effect of teacher proficiency for the SR task, $F(2, 156) = .08$, $p = .924$, $\eta_p^2 = .001$, or the PD task, $F(2, 156) = 1.06$, $p = .348$, $\eta_p^2 = .01$. A significant teacher proficiency \times test point interaction was found for the PD task, $F(3.49, 272.09) = 3.82$, $p = .007$, $\eta_p^2 = .05$, but not for the SR task, $F(4, 312) = 2.13$, $p = .08$, $\eta_p^2 = .03$.

Bonferroni tests for the PD task indicated that while learners taught by a primary school teacher with degree-level French made significant progress between all test points ($p < .001$, T1–T2 $d = 0.36$, T2–T3 $d = 0.46$, T1–T3 $d = 0.80$), those in the GCSE or lower teacher proficiency group only made significant progress between Years 5 and 6 ($p = .044$, $d = 0.66$). For those learners taught by native speakers, significant progress occurred only between Years 6 and 7 ($p < .001$, $d = 0.69$) and between Years 5 and 7 ($p < .001$, $d = 0.73$), but not between Years 5 and 6 ($p = .168$). At no test point was there any significant difference between scores across teacher proficiency groups.

Teacher Training and Global Scores

For teacher training, while there was a significant effect of test point for the SR task, $F(2, 314) = 18.13$, $p < .001$, $\eta_p^2 = .10$, and the PD task, $F(1.75, 274.90) = 31.95$, $p < .001$, $\eta_p^2 = .17$, there was no significant main effect of teacher training for either the SR task, $F(1, 157) = .17$, $p = .69$, $\eta_p^2 = .001$, or the PD task, $F(1, 157) = .13$, $p = .72$, $\eta_p^2 = .001$, and no significant two-way interaction for either the SR task, $F(2, 314) = 1.71$, $p = .18$, $\eta_p^2 = .01$, or the PD task, $F(1.75, 274.90) = 2.66$, $p = .08$, $\eta_p^2 = .02$. While the assumptions for homogeneity of variances and covariances were met for the SR test scores,

Levene's test was significant for PD scores in Year 5 ($p = .006$) and Year 7 ($p = .025$). The PD scores for all test points were, therefore, transformed using a square-root transformation and the analysis repeated. This confirmed the results of the original analysis, namely, a significant main effect for test point but not for teacher training and no significant two-way interaction.

Teacher Proficiency and PD Grammar

An analysis of PD grammar was then conducted; this was the area in which the highest correlation with teacher proficiency was observed in Year 7. We first explored the progress of learners in each teacher proficiency group. After a significant Friedman test, post hoc Wilcoxon signed ranks tests ($\alpha = .006$) indicated, as they had for global scores, that only groups taught at primary school by a teacher with degree-level French made significant progress in grammar across all test points, with small to medium effect sizes for T1–T2 ($z = 5.31, p < .001, r = .31$), for T2–T3 ($z = 5.62, p < .001, r = .23$), and for T1–T3 ($z = 3.22, p = .001, r = .39$). While those taught by a teacher with GCSE or lower made no significant progress at any point, the native speaker group did not make significant progress between Years 5 and 6 ($z = .202, p = .84$); however, they made rapid progress at the start of Year 7, nearly doubling their Year 6 score with a large effect size ($z = 4.81, p < .001, r = .52$).

Comparing PD grammar across teacher proficiency groups, Kruskal-Wallis tests showed that scores differed across groups at all test points ($p < .01$). Follow-up Mann-Whitney U tests (Table 6) showed no significant difference between the degree and GCSE teacher group at any point. By contrast, the learners in the native speaker teacher group outperformed the GCSE teacher group in Year 5 ($r = .31$), Year 6 ($r = .32$), and Year 7 ($r = .53$) and also significantly outperformed the degree teacher group in Years 5 ($r = .25$) and 7 ($r = .40$), but not in Year 6.

Teaching Time

Given that teaching time was the teaching variable most strongly correlated with outcomes (Table 5), we conducted a further repeated-measures ANOVA on global scores, with teaching time as a between-subjects factor and test point as a within-subjects factor. There was a significant main effect for test point in the SR task, $F(2, 312) = 68.42, p < .001, \eta_p^2 = .31$, and the PD task, $F(1.76, 273.86) = 69.10, p < .001, \eta_p^2 = .31$, as well as a significant main effect for teaching time in the SR task, $F(3, 156) = 6.73, p < .001, \eta_p^2 = .12$, and the PD task, $F(3, 156) = 7.88, p < .001, \eta_p^2 = .13$. There was also a significant two-way interaction in the SR task, $F(6, 312) = 3.19, p = .005, \eta_p^2 = .06$,

Table 6 Descriptive statistics and comparisons by teacher proficiency for grammar scores in the photo description task (Mann-Whitney U test, $\alpha = .006$)

Test point	GCSE or below			Degree			Native speaker		
	<i>Mdn</i>	<i>Range</i>	<i>n</i>	<i>Mdn</i>	<i>Range</i>	<i>n</i>	<i>Mdn</i>	<i>Range</i>	<i>n</i>
Year 5	3.00	0–12	49	3.00	0–16	146	6.00	0–18	57
Year 6	4.00	0–14	46	4.00	0–15	141	6.00	0–19	54
Year 7	4.00	0–11	16	5.00	0–16	105	11.00	2–24	43

Test point	GCSE vs. Degree		GCSE vs. Native speaker		Degree vs. Native speaker	
	<i>U</i>	<i>p</i>	<i>U</i>	<i>p</i>	<i>U</i>	<i>p</i>
Year 5	3,500.00	.820	902.50	.002*	2,718.50	.001*
Year 6	2,732.00	.107	779.50	.001*	3,020.50	.025
Year 7	619.00	.090	106.00	.001*	1,096.00	.001*

Note. GCSE = General Certificate of Secondary Education.

and the PD task, $F(5.27, 273.86) = 4.69, p < .001, \eta_p^2 = .08$. As homogeneity of variance was violated for the PD scores at Test Point 1 (Levene's test $p = .007$), the ANOVA was repeated using transformed scores, confirming the results of the original analysis, namely, a significant main effect for test point and teaching time, with a significant two-way interaction.

Bonferroni tests showed that for the SR task, the learners receiving 60 minutes of instruction had significantly higher scores than all other teaching time groups in Year 5 ($p = .05, d = 0.71$ – 1.24) and Year 6 ($p = .01, d = 0.75$ – 1.47). In Year 7, they still had significantly higher scores than the 15-minutes ($p = .04, d = 1.15$) and the 30-minutes groups ($p = .049, d = 0.55$), but not the 40-minutes group ($p = .34$), who were catching up. During the primary school years, there were no statistically significant differences between any of the groups receiving below 60 minutes of instruction. Looking at progress over time (see Figure 1), while the Year 5–6 progress of the 60-minutes group was statistically significant ($p < .001, d = 0.47$), it was not significant between Years 6 and 7 ($p = .24$). The 15-minutes group made very little progress between Years 5 and 6 ($p = 1.00$), while the 30-minutes group ($p < .001, d = 0.31$ – $.76$) and the 40-minutes group ($p < .001, d = 0.54$ – 1.21) made significant progress over all 3 years. Despite the slowing of progress for the 60-minutes group, at the start of Year 7, all groups receiving under an hour of instruction a week were still below, or barely reaching, the level of performance achieved by that group nearly two years earlier in Year 5.

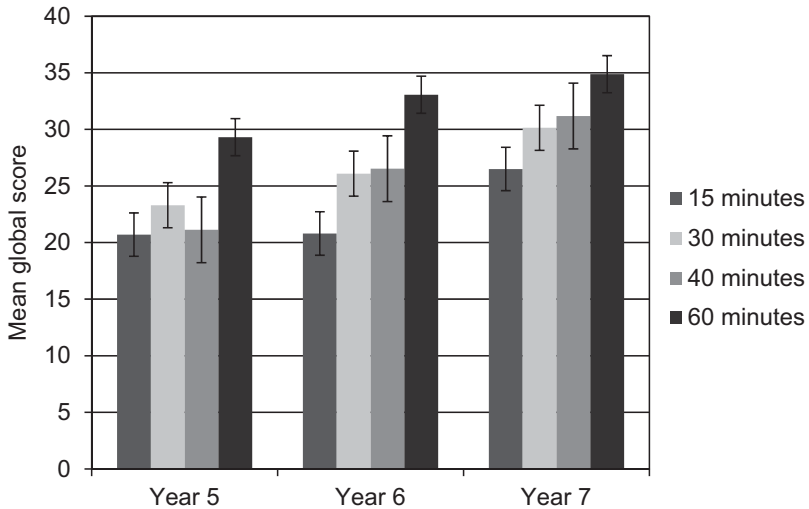


Figure 1 Mean global scores for the sentence repetition task by teaching time.

Similar results were found for Bonferroni tests conducted for the PD task. In Year 5, the 60-minutes learners again had significantly higher scores than learners in the 15-minutes ($p < .001$, $d = 1.78$), 30-minutes ($p < .001$, $d = 1.20$), and 40-minutes ($p < .001$, $d = 0.90$) groups. In Year 6, differences were still significant, but with smaller effect sizes: 60 versus 15 minutes ($p < .001$, $d = 1.61$), 60 versus 30 minutes ($p = .017$, $d = 0.64$), and 60 versus 40 minutes ($p = .007$, $d = 0.71$). At none of the test points were there any statistically significant differences between any of the groups receiving below 60 minutes of instruction. In terms of progress (see Figure 2), the 60-minutes group had a slightly lower Year 6 score, compared to Year 5, but there was renewed progress for them at the start of secondary school ($p = .003$, $d = 0.41$). Still, these learners did not significantly outperform any of the other groups once in secondary school. Between Years 5 and 6, the 15-minutes group made very little progress ($p = .27$), but moved forward significantly in Year 7 ($p < .001$, $d = 1.24$). Progress was made across all test points by the 30-minutes group ($p < .001$, $d = 0.38$ – 0.88) and the 40-minutes group ($p < .003$, $d = 0.29$ – 0.91).

Teaching Approach

The means for global scores displayed in Table 7 show little difference between the two approaches, with oracy learners seeming to have the greatest advantage on the PD task, but the difference was small. In order to further explore the

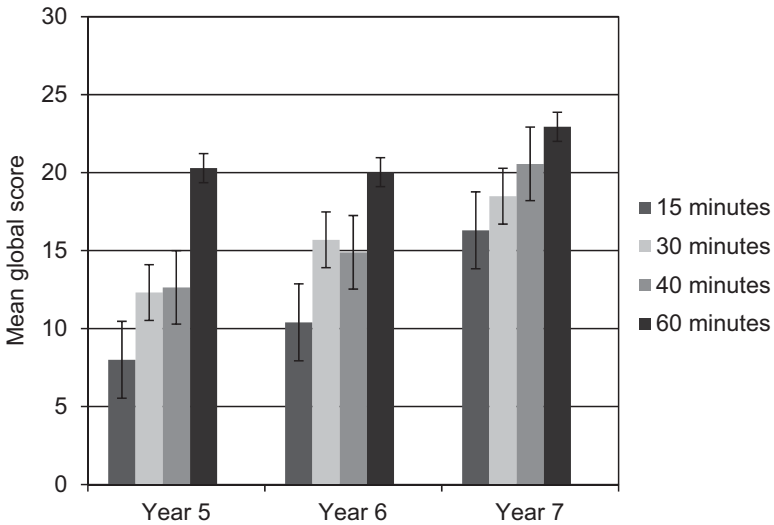


Figure 2 Mean global scores for the photo description task by teaching time.

Table 7 Means (standard deviations) for global scores by teaching approach in Years 5–7

Test point	Literacy			Oracy		
	<i>M</i>	<i>SD</i>	<i>Range</i>	<i>M</i>	<i>SD</i>	<i>Range</i>
Sentence repetition task						
Year 5	24.86	10.20	3–50	24.74	7.51	1–41
Year 6	28.17	10.74	6–49	27.85	8.69	7–45
Year 7	31.65	8.99	8–51	32.19	8.72	10–47
Photo description task						
Year 5	13.85	8.14	1–39	15.75	9.01	0–36
Year 6	16.04	8.06	2–36	17.31	8.41	0–40
Year 7	19.86	8.43	3–38	21.59	8.53	3–45

impact of teaching approach, a regression analysis was conducted in the GLM function of SPSS, permitting the inclusion of continuous and categorical factors (Hawkins, 2009). This enabled us to control for all primary school teaching variables, for which we entered the categorical variable school as a fixed factor alongside teaching approach. The school variable was used rather than teaching time, teacher proficiency, and teacher training separately, as the separate teaching variables were highly correlated with one another ($rho = .55-.74$).

Table 8 Means (standard deviations) for global scores for learners with low English literacy across teaching approaches

Test point	Literacy			Oracy		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Sentence repetition task						
Year 5	14.45	8.17	11	14.00	8.86	5
Year 6	17.30	7.75	10	15.75	7.14	4
Year 7	20.67	9.85	6	18.00	9.17	3
Photo description task						
Year 5	6.19	4.07	11	4.80	3.42	5
Year 6	8.50	6.57	10	4.25	6.33	4
Year 7	13.17	4.07	6	6.33	4.16	3

Learners' English literacy scores, which were significantly correlated with all outcomes ($\rho = .53-.63$), as discussed by Courtney, Graham, Tonkyn, and Marinis (2015), were entered as a covariate. The assumptions of homogeneity, normality, and linearity were checked and met (Hawkins, 2009). Regression models, summarized in Appendix S5 in the Supporting Information online, explained approximately 47–62% of the variance in scores across Years 5–7 for the SR task and between 51–57% for the PD task. While teaching approach did not have a significant impact on outcomes, school did, explaining 6–17% of the variance for the SR task, and 7–17% for the PD task in a subsequent set of regression analyses conducted with school used as the sole included variable. Thus, school became a more important explanatory factor as children moved into Year 7.

L1 Literacy

As previous literature suggested that a literacy-based approach might benefit learners with lower levels of L1 literacy, we also explored whether oracy and literacy approaches had different effects on these learners in our sample. Hence, we analyzed the scores of learners with a combined Year 5 English literacy score that was lower than one standard deviation below the total sample mean. This gave a very small number of learners (see Table 8), which reduced further as learners moved into secondary school. Mean global scores for the SR and PD tasks were calculated, with teaching approach as the grouping factor. Learners from both teaching approaches made similar progress across the three test points for the SR task, although the literacy group always outscored the oracy group. For the PD task, however, there was clearer evidence of progress between Years 6

and 7 for the learners receiving a literacy approach in primary school compared with those receiving an oracy approach. A Mann-Whitney U test, however, indicated that the difference was not statistically significant once a Bonferroni correction was applied ($Mdn_{\text{literacy}} = 12.50$, $Mdn_{\text{oracy}} = 5.00$, $U = 1.00$, $p = .048$, $\alpha = .017$).

Discussion

Our findings can be summarized as follows: Learners of French made statistically significant progress as they moved through Years 5, 6, and 7, but the amount of their progress was modest from one year to the next, particularly for grammar. This lends weight to the argument that early language learning in instructed contexts is not the rapid, effortless enterprise it is often assumed to be (e.g., as discussed by Muñoz, 2006). The fact that learners, on average, made progress across transition from primary to secondary school is contrary to what has been found in several studies (such as those from Europe summarized in Blondin et al., 1998), but in line with more recent investigation, including Courtney (2014), arguably because we tracked development in the same linguistic items longitudinally rather than using more open-ended tasks in which progress can be difficult to track. The factors of teaching time available for French at primary school and the primary school teachers' level of French proficiency and training in language teaching were all positively related to learning outcomes, although to differing degrees at different test points.

Length of Instruction

Correlations for teaching time were generally the strongest among all teaching variables, echoing the findings of previous research (e.g., Netten & Germain, 2008; Nikolov, 2009) regarding the importance of time for learning. Teaching time correlations were strongest in the primary years, perhaps indicating that time is especially important for younger learners, to enable them to draw on the more implicit mechanisms believed to underpin early language learning (DeKeyser, 2003; Muñoz, 2014). At primary school, only learners receiving 60 minutes of instruction significantly outperformed all other groups on all measures, suggesting that an hour a week is a threshold that needs to be reached before any difference in learners' progress related to amount of instruction can be detected. The lower correlations between outcomes and teaching time at the start of Year 7 may simply reflect the fact that all learners were receiving very similar amounts of teaching time by then. Furthermore, the slowing of progress for the 60-minutes group suggests that secondary schools were concentrating on bringing all learners within a heterogeneous intake to a similar level of

proficiency, possibly at the expense of building on the greater progress made by the 60-minute learners at primary school.

Teachers' Proficiency and Training

By contrast, the primary school teachers' level of training and their language proficiency were more strongly related to outcomes in Year 7 than in earlier years. The relationship between teacher training and outcomes reinforces the findings of studies in other contexts (e.g., Edelenbos & Suhre, 1994; Szpotowicz, 2009), but with regard to learners' progress over time, teachers' language proficiency emerged as more important in the present study than their training. This was particularly true for grammar on the PD task, supporting the argument that learners need an adequate model, with input that has consistency of grammatical features in order to acquire them (Murphy, 2014). While progress was most even across the three test points for learners taught by a teacher with degree-level French, groups taught by native speakers showed little progress between Years 5 and 6, but then large amounts in Year 7. Although we did not directly measure the quality and quantity of teacher target language input in our schools, it is possible that native speaker teachers provided the richer input that Aukrust (2007) argues may initiate processes that develop in the longer term rather than immediately. The more consistent progress made by groups that were taught by nonnative teachers with a degree in French may have resulted from the teachers providing a good model and a more scaffolded approach, as some research suggests is associated with nonnative teachers (e.g., Walkinshaw & Duong, 2012).

Literacy Versus Oracy

The teaching approach learners received, whether oracy or literacy, did not seem to influence their levels of attainment when learners of different English literacy levels were considered together. This finding reflects the conclusions of Edelenbos and Suhre (1994), and those reported in Blondin et al. (1998) for a large body of European research, that course is less important than other classroom factors. The only tentative benefit found for teaching approach was that learners with lower levels of English language literacy seemed to make more progress as they moved into secondary school and in relation to production if they had received French instruction with a stronger literacy focus at primary school. This indicates, perhaps, that such learners need access to the written form in order to facilitate retention of vocabulary and grammar forms, as suggested by Hu (2008) in relation to learners with lower levels of phonological awareness and by Harley and Swain (1984) commenting on the needs of young

learners. It is possible that presenting language in an oral form only places a heavy burden on such learners, preventing them from retaining the language to which they are exposed or from recalling it easily.

The lack of a clearer impact of literacy approach on outcomes across all learners, however, suggests more strongly that quantity and quality of input were more important than focus of instruction, as indicated by the results of the regression analysis in which teaching approach had no impact on outcomes once the combined variable of school (including teaching time, teacher training, and teacher French proficiency) was entered into the model. It is also possible, however, that the differences between the oracy and literacy approaches used by teachers were not clear enough, even after we had excluded borderline schools. In Year 6, the amount of literacy activities increased across all schools, as teachers prepared learners for the more literacy-focused approach they would meet at secondary school.

Furthermore, even in schools receiving the highest literacy score (7 and 8 out of a possible 10), there was relatively little evidence of the higher-level literacy activities that may be needed for literacy to support learning. In only one school did we see such activities, where learners read a short text and answered questions about it and also wrote sentences based on a model. Perhaps not coincidentally, learners in that school (School 1) had the highest mean scores when individual school scores were examined. It is possible that with more focus on higher-level literacy activities, a clearer advantage for literacy schools might have been found. However, learners in School 1 also experienced the greatest amount of teaching time (60 minutes, like one other oracy school); arguably, sufficient lesson time is needed in order to incorporate activities that go beyond word and sentence level. Considering this question from another angle, it might be that the teacher of School 1 learners was able to include literacy activities precisely because her learners had made sufficient progress through higher-quality teaching that enabled them to cope with such activities.⁶

The findings from School 1 thus suggest that teaching approaches and other teacher and teaching factors interact in complex ways. For example, the two schools that allocated the most time to the teaching of French (Schools 1 and 7) both employed teachers with a degree or native speaker competence in French and who had completed specialist postgraduate teacher training in French teaching. One teacher was in the oracy group, the other in the literacy group. By contrast, the teacher in the school allocating the least amount of time to French per week (15 minutes) had had relatively little training in teaching it, although she had a degree in the subject. This suggests, perhaps, that primary schools, where French is seen as important, ensure that it is taught

by a well-trained, linguistically proficient teacher, and allocate sufficient time to its teaching. The growing amount of variance explained by the school variable across our test points suggests that it is the coming together of a number of factors that is important for the best possible outcomes for young learners as they move from primary into secondary school, rather than one single factor that makes the difference.

Limitations

Levels of attrition in our sample size in Year 7 pose a limitation to our exploration of the impact of teaching and teacher factors, as does the nature of the literacy–oracy division we adopted. Future research into the impact of teaching approach would most likely benefit from a strictly experimental perspective, with the teaching at any schools manipulated in a tightly controlled manner, including controlling for teaching time available and teacher language proficiency and training. In the present study, we sought instead to take a more naturalistic approach, to study practice that is actually occurring in schools. While this approach inevitably has its shortcomings, it can also be argued that it gives a more grounded picture of early language learning classrooms.

Our findings suggest, albeit tentatively, that a literacy-based approach has the potential to help children with lower levels of L1 literacy make progress in learning another language, but these findings are limited by the small sample of such learners that we had. Further research targeting these learners is warranted, given that they may be excluded from language classrooms, as occurs, for example, in the United States (Sparks, 2012) and increasingly in England, especially at secondary school, on the grounds that they would find language learning too difficult and achieve limited progress (Tinsley & Board, 2016). Such research would also benefit, arguably, from a consideration of the role of working memory capacity in relation to different types of instruction, an area we were unable to explore in the present study. It should also be acknowledged that we only assessed learners' grammar and vocabulary knowledge orally and that learners had already experienced 3 years of learning French prior to the start of the study. While it is probable that learners received similar amounts of teaching time prior to Years 5 and 6, we cannot be certain of this, nor of the other teaching and teacher factors to which they were exposed during periods prior to the study.

Conclusion

The significance of this study lies in the evidence it provided from a longitudinal investigation, with a relatively large sample, regarding the importance of

optimizing the conditions for learning a foreign language at primary school to provide more equal opportunities for all learners, namely, through teachers with sufficient pedagogical and linguistic expertise and sufficient teaching time. Such conditions are far from being guaranteed either in England (Tinsley & Board, 2016) or elsewhere (Copland et al., 2014). Thus, the findings of the study are of relevance beyond the context of England, across countries implementing a lowering of the age at which the learning of another language begins. Such policies imply that language learning is a valued part of the curriculum. If this is the case, then as much if not more attention needs to be paid to providing the minimal conditions for learning—in terms of teaching time, teachers' expertise and training, and teaching approach—compared to the attention devoted to learners' age. As expressed by Muñoz (2011), “trusting young age of learning with the burden of learning success is clearly not enough” (p. 130). We very much concur with that view.

Final revised version accepted 14 April 2017

Notes

- 1 We thank an anonymous reviewer for this suggestion.
- 2 We are aware that the assessment of grammatical gender can be problematic in that learners have a 50–50 chance of assigning the correct gender (masculine or feminine). To take account of the possibility of guessing, we examined responses for grammatical gender separately using a one-sample *t* test. At each test point, learners' performance on gender was significantly different than chance ($p < .05$).
- 3 All teachers were on the primary schools' core staff and most were employed specifically to teach French. In three cases, French was taught by a generalist teacher who taught them all other curriculum subjects.
- 4 Effect size values were interpreted as follows (Field, 2013): $d = 0.2$ (small), $d = 0.5$ (medium), $d = 0.8$ (large); $r = .10$ (small), $r = .30$ (medium), $r = .50$ (large).
- 5 It is acknowledged that groups remained rather unequal in size.
- 6 This was suggested by an anonymous reviewer.

References

- Aukrust, V. G. (2007). Young children acquiring second language vocabulary in preschool group-time: Does amount, diversity, and discourse complexity of teacher talk matter? *Journal of Research in Childhood Education*, 22, 17–37.
<https://doi.org/10.1080/02568540709594610>
- Blondin, C., Calendier, M., Edelenbos, P., Johnstone, R., Kubanek-German, A., & Taeschner, T. (1998). *Foreign languages in primary and pre-school education: A review of recent research within the European Union*. London: CILT.

- Bowers, E. P., & Vasilyeva, M. (2011). The relation between teacher input and lexical growth of pre-schoolers. *Applied Psycholinguistics*, 32, 221–241.
<https://doi.org/10.1017/S0142716410000354>
- Cable, C., Driscoll, P., Mitchell, R., Sing, S., Cremin, T., Earl, J., et al. (2010). *Language learning at Key Stage 2: A longitudinal study*. London: Department for Children, Schools and Families (DCSF Research Reports DCSF-RR198).
- Campfield, D. E., & Murphy, V. A. (2014). Elicited imitation in search of the influence of linguistic rhythm on child L2 acquisition. *System*, 42, 207–219.
<https://doi.org/10.1016/j.system.2013.12.002>
- Copland, F., Garton, S., & Burns, A. (2014). Challenges in teaching English to young learners: Global perspectives and local realities. *TESOL Quarterly*, 48, 738–762.
<https://doi.org/10.1002/tesq.148>
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.
- Courtney, L. (2014). *Moving from primary to secondary education: An investigation into the effect of primary to secondary transition on motivation for language learning and foreign language proficiency*. Unpublished doctoral dissertation, University of Southampton, UK.
- Courtney, L., Graham, S., Tonkyn, A., & Marinis, T. (2015). Individual differences in early language learning: A study of English learners of French. *Applied Linguistics*.
<https://doi.org/10.1093/applin/amv071>
- DeKeyser, R. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 313–348). London: Blackwell.
- Department for Education. (2005). *Key stage 2 framework for languages*. Nottingham, UK: DfES Publications.
- Department for Education. (2013). *Languages programmes of study: Key stage 3. National Curriculum in England*. Retrieved 12 January 2017 from
<https://www.gov.uk/government/publications/national-curriculum-in-england-languages-programmes-of-study>
- Devescovi, A., & Caselli, C. (2007). Sentence repetition as a measure of early grammatical development in Italian. *International Journal of Language and Communication Disorders*, 42, 187–208.
<https://doi.org/10.1080/13682820601030686>
- Drew, I. (2009). Using the Early Years Literacy Programme in primary EFL Norwegian classrooms. In M. Nikolov (Ed.), *Early learning of modern foreign languages: Processes and outcomes* (pp. 108–120). Clevedon, UK: Multilingual Matters.
- Driscoll, P., Martin, C., Graham-Matheson, L., Dismore, H., & Sykes, R. (2004). *A systematic review of the characteristics of effective foreign language teaching to pupils between the ages 7 and 11*. Research Evidence in Education Library.

- EPPI-Centre, Social Science Research Unit, Institute of Education,
London.
- Edelenbos, P., & Suhre, C. (1994). A comparison of courses for English in primary education. *Studies in Educational Evaluation*, 20, 513–534.
- Enever, J. (Ed.). (2011a). *Early language learning in Europe*. London: British Council.
- Enever, J. (2011b). Policy. In J. Enever (Ed.), *Early language learning in Europe* (pp. 23–32). London: British Council.
- Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, 27, 464–491.
<https://doi.org/10.1093/applin/aml001>
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*. London: Sage.
- Galton, M., Gray, J., & Ruddock, J. (1999). *The impact of school transitions and transfers on pupil progress and attainment*. DfEE Research Report No. 131. Norwich, UK: HMSO.
- Genelot, S. (1997). L'enseignement des langues vivantes à l'école élémentaire: Eléments d'évaluation des effets au collège. *Revue Française de Pédagogie*, 118, 27–42.
- Harley, B., & Swain, M. (1984). The interlanguage of immersion students and its implications for second language teaching. In A. Davies, C. Criper, & A. P. R. Howatt (Eds.), *Interlanguage* (pp. 291–311). Edinburgh, UK: Edinburgh University Press.
- Hawkins, D. (2009). *Biomeasurement*. Oxford, UK: Oxford University Press.
- Hill, K., Davies, A., Oldfield, J., & Watson, N. (1998). Questioning an early start: The transition from primary to secondary foreign language learning. *Melbourne Papers in Language Testing*, 6, 21–36.
- Hu, C. F. (2008). Use orthography in L2 auditory word learning: Who benefits? *Reading and Writing*, 21, 823–841. <https://doi.org/10.1007/s11145-007-9094-6>
- Jaekel, N., Schurig, M., Florian, M., & Ritter, M. (2017). From early starters to late finishers? A longitudinal study of early foreign language learning in school. *Language Learning*. <https://doi.org/10.1111/lang.12242>
- Jessop, L., Suzuki, W., & Tomita, Y. (2007). Elicited imitation in second language acquisition research. *Canadian Modern Language Review*, 64, 215–220.
<https://doi.org/10.3138/cmlr.64.1.215>
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford, UK: Pergamon Press.
- Lightbown, P. M. (2014). Making the minutes count in L2 teaching. *Language Awareness*, 23, 3–23. <https://doi.org/10.1080/09658416.2013.863903>
- Lo Bianco, J. (2009). *Second languages and Australian schooling*. Melbourne, Australia: Australian Council for Educational Research.
- Low, L., Brown, S., Johnstone, R., & Pirrie, A. (1995). *Foreign languages in primary schools: Evaluation of the Scottish pilot projects 1993–5*. Final report to Scottish Office. Stirling, UK: Scottish CILT.

- Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 95–124). Clevedon, UK: Multilingual Matters.
- Mihaljević Djigunović, J. (2009). Impact of learning conditions on young FL learners' motivation. In M. Nikolov (Ed.), *Early learning of modern foreign languages: Processes and outcomes* (pp. 75–89). Clevedon, UK: Multilingual Matters.
- Muñoz, C. (2006). The effects of age on foreign language learning: The BAF project. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 1–40). Clevedon, UK: Multilingual Matters.
- Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, 29, 578–596.
<https://doi.org/10.1093/applin/amm056>
- Muñoz, C. (2011). Is input more significant than starting age in foreign language acquisition? *International Review of Applied Linguistics*, 49, 113–133.
<https://doi.org/10.1515/iral.2011.006>
- Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, 35, 463–482.
<https://doi.org/10.1093/applin/amu024>
- Murphy, V. (2014). *Second language learning in the early school years: Trends and contexts*. Oxford, UK: Oxford University Press.
- Myles, F., & Mitchell, R. (2011). *Learning French from ages 5, 7 and 11: An investigation into starting ages, rates and routes of learning amongst early foreign language learners*. ESRC End of Award Report, RES-062-23-1545. Swindon, UK: ESRC.
- Netten, J., & Germain, C. (2008). The future of intensive French in Canada. *Canadian Modern Language Review*, 65, 757–786. <https://doi.org/10.1353/cml.0.0077>
- Nikolov, M. (2009). Early modern foreign language programmes and outcomes: Factors contributing to Hungarian learners' proficiency. In M. Nikolov (Ed.), *Early learning of modern foreign languages: Processes and outcomes* (pp. 90–107). Clevedon, UK: Multilingual Matters.
- Office for Standards in Education. (2011). *Modern languages: Achievement and challenge 2007–2010*. Retrieved 12 January 2017 from <http://www.ofsted.gov.uk/resources/modern-languages-achievement-and-challenge-2007-2010>
- Porter, A. (2014). *An early start to French literacy: Learning the spoken and written word simultaneously in English primary schools*. Unpublished doctoral dissertation, University of Southampton, UK.
- Pufahl, I., & Rhodes, N. C. (2011). Foreign language instruction in U.S. schools: Results of a national survey of elementary and secondary schools. *Foreign Language Annals*, 44, 258–288. <https://doi.org/10.1111/j.1944-9720.2011.01130.x>

- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin.
- Sparks, R. L. (2012). Individual differences in L2 learning and long-term L1-L2 relationships. *Language Learning*, 62, 5–27.
<https://doi.org/10.1111/j.1467-9922.2012.00704.x>
- Szpotowicz, M. (2009). Factors influencing young learners' vocabulary acquisition. In M. Nikolov (Ed.), *Early learning of modern foreign languages: Processes and outcomes* (pp. 195–211). Clevedon, UK: Multilingual Matters.
- Tinsley, T., & Board, K. (2016). *Language trends 2015/16: The state of language learning in primary and secondary schools in England*. Reading, UK: CfBT. Retrieved 12 January 2017 from
https://www.britishcouncil.org/sites/default/files/language_trends_survey_2015.pdf
- Unsworth, S., Persson, L., Prins, T., & De Bot, K. (2015). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics*, 36, 527–548. <https://doi.org/10.1093/applin/amt052>
- Walkinshaw, I., & Duong, O. T. H. (2012). Native- and non-native speaking English teachers in Vietnam: Weighing the benefits. *TESL-EJ*, 16. Retrieved 12 January 2017 from <http://www.tesl-ej.org/wordpress/issues/volume16/ej63/ej63a1>
- Wu, S.-L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals*, 46, 680–704.
<https://doi.org/10.1111/flan.12063>
- Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*, 33, 497–528. <https://doi.org/10.1177/0265532215594643>

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Teacher and Teaching Factors in Each Primary School.

Appendix S2. Description and Scoring of Sentence Repetition Task.

Appendix S3. Description and Scoring of Photo Description Task.

Appendix S4. Mean Global Scores Across Teacher Groups.

Appendix S5. GLM Regression Analysis for Sentence Repetition and Photo Description Tasks.