Majority language vocabulary and nonword repetition skills in children attending minority language immersion education

Abstract
The present study examined nonword repetition (NWR) and comprehension/production of single word vocabulary in the majority language (English) in six- to eight-year-old English-Gaelic emergent bilingual children attending Gaelic-medium primary education (GMPE) (primary years 2 and 3). GMPE is an immersion education model found in Scotland where the minority language, Gaelic, is the language of instruction, whereas English is the majority community language, not supported in school in the first three years of primary schooling. All children spoke English as their first language (L1) and had varying exposure to the second language (L2) Gaelic outside the school setting. We investigated how individual factors such as age and frequency of exposure, and psycholinguistic factors, such as task modality, word class (for vocabulary), frequency, length, and prosody (for nonword repetition), influenced performance. Later exposure to Gaelic and higher nonverbal IQ were contributing predictors of performance for English vocabulary. For NWR, participants with earlier and longer exposure to Gaelic who also used both languages at home more performed significantly better on this task. These results suggests that minority language immersion education differentially affects majority language abilities; later L2 onset benefited L1 vocabulary learning, whereas earlier L2 onset boosted L1 NWR skills in these bilinguals.
Introduction

Whereas education in a majority second language (L2) is a necessity for children who speak a minority language (e.g., children from an immigrant background growing up in a country where the minority language is not the language of education or the wider community), for children speaking a majority language (e.g., English in the UK), attending educational programmes in an L2 that is a minority language in the community, is an educational choice. Generalising across various immersion education programmes, children in minority language immersion programmes constitute a distinct type of bilingual children not least because teaching of the majority language within the school context is usually suspended, especially in the early years (usually the first three years of primary school), while minority language use is prioritised to promote its revitalisation and maintenance (Paradis, Genesee, & Crago, 2011).

A question that arises for majority language children attending minority language immersion education is how the majority language that is unsupported during the early primary school years develops during that time. A breadth of studies across different countries seems to suggest that majority language children who participate in immersion education acquire typical levels of native language development by the end of primary schooling (Bialystok, 2018; Paradis et al., 2011 for overviews). However, (at least) two aspects of bilingual development remain particularly less well understood in an immersion education context. First, most studies focusing on the comparison between majority language students in immersion education and their monolingual peers in mainstream education have used standardised language assessments. This does not allow us to investigate the psycholinguistic and individual factors that may affect language development in majority language children in minority immersion education. This is particularly important for emergent bilingual children in the early years, when the majority language is not supported
yet in the school setting (Montanari, 2014; Padilla, Fan, Xu, & Silva, 2013; Paradis et al., 2011). Second, vocabulary and nonword repetition, albeit related, have been shown to be 

differentially 

affected by bilingualism-related factors, such as the age (AoE) and length of exposure (LoE) to a particular language (Chondrogianni, 2018; Golberg, Paradis, & Crago, 2008; Paradis, 2011). Although both factors influence language outcomes in bilinguals, vocabulary has been shown to be more susceptible to exposure effects than nonword repetition for sequential bilingual children, when tested in their weaker language (Chondrogianni & Marinis, 2011; Oller, Pearson, & Cobo-Lewis, 2007; Paradis et al., 2011). How exposure factors affect majority language outcomes in emergent bilingual children attending minority L2 immersion education remains controversial with mixed results (V. C. Gathercole & Thomas, 2009; Rhys & Thomas, 2013).

Vocabulary and nonword repetition are not only related (S. E. Gathercole et al., 1992; Masoura & Gathercole, 1999) but are also good predictors of academic ability (S. E. Gathercole et al., 1992). Vocabulary knowledge is pivotal for young children’s language development and an important predictor of later academic skills, in particular early literacy and reading (Duff et al., 2015). Nonword repetition (NWR) as an index of phonological short-term memory (STM) (S. E. Gathercole & Baddeley, 1990) but also of sublexical representations, such as phonotactics and syllable structure (Jones et al., 2007), refers to the individual’s ability to temporarily retain phonemes in the working memory in the correct order while creating a motor plan to articulate auditory information. This ability is not only predictive of the individual’s ability to learn new words (S. E. Gathercole et al., 1992), but also the more words an individual knows, the better they are at producing nonwords by making use of lexical and sublexical information (Szewczyk et al., 2018). For that reason, nonword repetition correlates with various aspects of grammatical and vocabulary
development, which in turn predict academic attainment (e.g., Gathercole, Willis, Emslie, & Baddeley, 1992).

In studies with monolingual children and, increasingly, with bilingual children, assessing vocabulary comprehension and production in the majority first (L1) or second language (L2) in mainstream education, psycholinguistic factors such as the class a word belongs to (noun or verb), or when it is estimated to be acquired, also known as the subjective estimation of the age of acquisition of a word (AoA), have been shown to influence language development (Haman et al., 2015; Hansen et al., 2017; van Wonderen & Unsworth, 2020). Similarly, monolingual and bilingual children’s ability to correctly repeat nonwords is affected by how many phonological neighbours a nonword has, and hence how likely it is to be phonologically similar to a real word, its prosody and phonotactic probability (Chiat, 2015; Chiat & Polišenská, 2016a; Szewczyk et al., 2018). However, in existing studies the focus has been on the majority L1 (for monolingual children) or L2 (for sequential bilingual children) which are supported within the school context. Whether and how these factors interact in the same way and to the same degree in the majority L1 of bilingual children attending L2 minority immersion education remains open to investigation.

In Scotland, immersion education has been instantiated in the form of Gaelic medium education (GME). GME is a type of immersion education that aims to revitalise Scottish Gaelic (Gaelic from this point onwards), one of the heritage languages of the country, by increasing its use among younger generations through Gaelic immersion education. In the present study, we investigated whether individual factors, such as the child’s current age, and exposure factors, such as the age when they started learning Gaelic and their degree of exposure to the two languages affect the development of the majority L1 (English) vocabulary and nonword repetition skills of pupils attending minority L2 (Gaelic) education in the early primary school years. We also investigated how psycholinguistic, item-level
factors, such as the class (noun vs verb) a word belongs to and its estimated AoA (for vocabulary), the number of syllables, phonotactic structure, and prosody of a nonword affect performance on vocabulary and nonword repetition, respectively. By focusing on the early years of immersion education, what is called “early total immersion” (Genesee, 2006), we were able to investigate how these language and academic-related skills develop in majority language pupils and how they are modulated by child- and item-specific factors in these emergent bilingual children (Paradis et al., 2011).

**Majority language abilities of children in immersion education**

There is a broad consensus in the literature that the majority language skills of children in minority language immersion education are generally on a par with the language skills of children in mainstream majority language education by the time they reach later primary school years (Bialystok, 2018; Paradis et al., 2011 for reviews of relevant studies). This conclusion seems to hold regardless of cultural context or linguistic distance between the languages in immersion programmes (Montanari, 2013; Padilla et al., 2013; Paradis et al., 2011). This is attributed to three main reasons. First, the exposure that majority language students receive in the community dominant language seems to offset any effects of lack of educational support in immersion education, at least in the early years of immersion education (Paradis et al., 2011). Second, the support the majority language receives in the later years of immersion education (e.g., after year 3 in primary education) compensates for any early lag (Genesee, 2006). Finally, academic and literacy skills developed in one language, even the minority L2, seem to transfer rather quickly and boost performance in the majority language (Cummins, 2000).

However, the bilingualism related variables that may influence majority language outcomes in immersion education, such age and length of exposure to the minority language
through immersion education, are less well understood. The timing of L2 immersion has been shown to affect language outcomes in both the majority and the minority language. Studies with L1 English - L2 French-speaking children in the early years of French-Canadian immersion education have shown that those with an early L2 French immersion lag behind their monolingual English-speaking peers in mainstream majority language education on their L1 English performance on standardised assessments targeting literacy and mathematics compared to children with later L2 immersion, although performance improves with age (Genesee, 2006). Conversely, minority language immersion in the preschool and early primary school years leads to better performance in the minority language compared to late immersion, e.g., at the beginning of secondary school (Paradis, Genesee, & Crago, 2011), including immersion education for revitalisation, such as in the case of Irish (Ó Muircheartaigh & Hickey, 2008).

In Scotland, GME is an immersion model with its main aim the revitalisation of Gaelic, a minority language of the Celtic language family with an official status, since the Gaelic Language Act was introduced by the Scottish government in 2005, but under immense pressure from the majority, societal language, English. According to the most recent Census (2011), only 1.1% of the Scottish population can speak Gaelic (approx. 58,000 speakers out of the total population of 87,000 who are able to speak/read/understand Gaelic), but there was an increase between 2001 and 2011 in the proportion of young people who speak the language (National Records of Scotland, 2015, p. 9), mostly boosted by GME provision. GME is an immersion model found in Scotland that spans across preschool, primary, and secondary education and targets the acquisition of both Gaelic and English with the view to make children fully bilingual by the time they enter secondary education. Gaelic is prioritised in the first three years (P1-P3) of Gaelic-medium primary education (GMPE) and English is introduced slowly in lessons from P4 (O’Hanlon, Paterson and McLeod 2012). Pupils
entering GMPE come from a variety of backgrounds. Most pupils come from families with no Gaelic at home and are immersed in Gaelic at school (Stephen et al. 2010). A small proportion of parents (approximately 18%) are native speakers of Gaelic (O’Hanlon et al. 2012). Nursery provision may or may not be attached to school(s) in areas that offer GMPE, and when it is, it may not be offered full-time. As a result, pupils enter primary schools with mixed prior experience of formal instruction in Gaelic. The combination of its minority status along with its use in primarily educational contexts places GME more on a par with other immersion programmes where language revitalisation through education is the main vehicle of this effort (as in some indigenous language immersion programmes (Comajoan-Colomé & Coronel-Molina, 2021)), rather than immersion programmes where the minority language is a majority language elsewhere, and potentially spoken by millions of speakers (Baker, 2010). An example of the latter would be the case of French-immersion programmes in Canada, with French being considered a minority in some Canadian provinces, but it is spoken by millions of people in France and globally.

Two studies to date have investigated the majority language (English) abilities of pupils in GMPE (Johnstone et al., 1999; O’Hanlon et al., 2013). Both studies investigated children in mid-primary school grades (P5-P7), that is children aged between 9 and 11 years old, using results from state-wide standardised assessments on science, mathematics, and English literacy abilities. Their findings converge on that pupils in GMPE performed similarly to (mathematics or science), or in some cases outperformed (reading) their peers in English mainstream education. These results are in line with existing findings from the international literature on immersion vs. monolingual education provision, which suggest that children in later years in immersion education achieve comparable performance in their majority language (e.g., see Paradis et al., 2011).
Given that both studies, however, focused on children in mid-primary education (children aged between 9 and 11 years old), we currently know little about how specific language areas of children in GMPE develop, and which individual factors modulate this development in the early years of primary school, that is, the first three years of schooling. To address this gap, in the present study, we tested children in what is called the first level of primary school in Scotland, namely P2 and P3 (aged between 6 and 8 years old), to examine how children’s majority language abilities develop in the early primary years, before the majority language is introduced into the curriculum, and once they have had at least one year of instruction in GMPE. We also evaluated how frequency and exposure to English and Gaelic outside the school context to Gaelic currently and before they started primary school affected performance on two specific linguistic domains that are predictors of linguistic and educational attainment, that is vocabulary knowledge and NWR. In the next section, we review how degree, length, and age of exposure to the two languages has been shown to influence the two language areas examined in the present study, namely vocabulary and NWR.

**Linguistic and item-specific variables affecting vocabulary and NWR in bilingual children**

Vocabulary knowledge encompasses the words the child can comprehend (receptive vocabulary) and produce (expressive vocabulary). Previous studies with children in majority language mainstream education have shown that both monolingual and L1 minority – L2 majority bilingual children performed better on receptive (comprehension) than expressive (production) tasks, but bilinguals had a larger gap between the two tasks compared to monolingual children (Gibson et al., 2012; Kaushanskaya & Marian, 2009; Oller et al., 2007). Children’s better performance on receptive compared with expressive tasks has been
attributed to the different psycholinguistic processes involved in comprehending versus producing a word (Levelt, 2001; Yan & Nicoladis, 2009). Whereas accessing semantic or phonological representations may be sufficient for success on receptive tasks, in an expressive task, participants are not only required to retrieve the correct lexical item but also articulate it, rendering production more demanding than comprehension (Levelt, 2001; Yan & Nicoladis, 2009). In the case of sequential bilingual children assessed in their L2, the receptive-expressive discrepancy has been attributed to the weak or underspecified semantic or phonological representations that may characterise bilingual children’s lexicon due to reduced language exposure, also known as the weaker links hypothesis (Gollan et al., 2008). Reduced language exposure may impact on how strong the links are between lexical access and semantic or phonological representations. This, in turn, may impact on bilingual children’s ability to access words in the lexicon during comprehension and, subsequently, produce them (Gibson et al., 2020). When exposure to a language is reduced, as is possible in a bilingual context, the discrepancy between receptive and expressive abilities may also increase (Gibson et al., 2020).

Knowing a word also entails knowing the word class it belongs to, i.e., whether a word is a noun or a verb. Nouns have been reported as earlier acquired compared to verbs (Haman et al., 2015; Luniewska et al., 2019) and children learn nouns more easily and have larger vocabulary sizes for nouns than verbs (Altman et al., 2017; Haman, Wodniecka, et al., 2017). In tasks usually administered to pre-school or school aged children, nouns depict concrete and imageable objects, which are perceptually easier to learn and identify than verbs, which require understanding events or actions and a relationship between participating entities (McDonough et al., 2011).

Vocabulary learning is also influenced by the age at which a word is acquired (AoA). Although AoA can be measured in different ways by looking for example at child corpora
(Juhasz, 2005), AoA can also be estimated using subjective, native speakers’ reports of when they thought they learnt a particular word (Haman et al., 2015). This measure, despite its subjectivity, has been found to consistently and reliably correlate with parental reports on young children’s lexical development across speakers and languages from a wide typological spectrum, regardless of their majority or minority status (Haman et al., 2015; Łuniewska et al., 2019). This is because subjective AoA indirectly measures how long a child has been using a word and how entrenched the use of a word might be. The earlier a word is acquired the more experience the speaker has with using this word in various contexts and the easier it is to access it and entails that a child has had more time to practice with (comprehend and produce) a particular word. This may explain why AoA has been found to very significantly predict accuracy on nouns and verbs in expressive and receptive vocabulary tasks similar to the ones used in the present study with bilingual children, with the earlier the AoA, the higher the accuracy on the word (Altman et al., 2017; van Wonderen & Unsworth, 2020). However, Haman et al. (2017) found no significant correlations between subjective AoA and expressive vocabulary in the 5- to 6-year-old British English monolingual children in their study. AoA only correlated with expressive vocabulary skills with verbs but not with nouns. This may have been because the monolingual children in their study had ceiling accuracy on the comprehension tasks.

NWR tasks require the participant to store, manipulate and repeat phonological information of items that are not real words and which participants have not previously heard. As such, they assess phonological STM on top of phonological processing and articulatory skills (S. E. Gathercole & Baddeley, 1990). At the same time, NWR tasks tap into an individual’s word knowledge, as well as knowledge of the language specific (LS), sublexical properties of words, such as word syllable structure, phonotactics, and prosody (stress) patterns (Chiat, 2015; Szewczyk et al., 2018). Because repetition of nonwords requires
children to learn a new, yet made-up, lexical item, factors that affect how children learn native words may affect performance on the NWR task. Children learn language on many levels, from statistical information about frequency of phonemes or combinations of phonemes, to whole words and syntax (Saffran & Wilson, 2003; Sundara et al., 2008). Specifically, sublexical representations containing phoneme combinations occurring more or less frequently in a language may support children’s nonword repetition skills (S. E. Gathercole et al., 1992; Masoura & Gathercole, 1999). This level of phonological knowledge has been proposed by theories of phonological knowledge, such as the EPAM-VOC and CLASSIC, that assume that the presentation of nonwords activates sequences of phonemes in the long-term memory that are present in the nonword (e.g., see (Jones et al., 2007). This explain why young children are sensitive to a sublexical property, such as the probability with which syllables within a word co-occur, known as transitional probability, and why this helps to identify words from a stream of spoken language (Cutler & Carter, 1987; Thiessen & Saffran, 2003), including in bilingual contexts (Sundara et al., 2008). Ngram frequency is another sublexical index similar to transitional probability but exceeds transitional probability in that it reflects how often phonemes, not just syllables, are found together in different-sized combinations in real words (Szewczyk et al., 2018). Because phonemic Ngrams measure phonotactic probability taking into consideration the frequency of phoneme sequences at many different lengths, they are much better at capturing subtle lexical and sublexical properties. For that reason, Ngram frequency has been shown to be a strong predictor for higher NWR accuracy beyond more qualitative ratings, such as word-likeness (Munson et al., 2005) and when consonant cluster complexity is reduced (Szewczyk et al., 2018). Consonant cluster complexity is another phonological factor that has been shown to determine children’s performance on NWR tasks, with nonwords containing more consonant clusters leading to lower accuracy than nonwords with fewer consonant clusters (Marshall & van der Lely,
This is because consonant clusters are phonologically complex, and producing consonant clusters requires a range of complex and rapidly changing oromotor movement than Consonant-Vowel (CV) sequences (MacNeilage & Davis, 2000).

Prosody is a sublexical property that refers to language-specific stress patterns and has been shown to affect word learning and also performance on NWR tasks, with children preferring both words and nonwords with language-specific prosody (Chiat, 2015). For example, children acquiring English show a preference for words and nonwords with trochaic foot structure, in which a stressed (strong) syllable is followed by an unstressed (weak) syllable (SW), a prosodic pattern found in 90% of bisyllabic English words (Roy & Chiat, 2004). In a study targeting prosody, five- to eight-year-old monolingual children performed better on a task with typical prosody of the target language than on a task with evenly-stressed non-words (Archibald et al., 2009; Archibald & Gathercole, 2007). Furthermore, children had higher whole-item accuracy and syllable-number accuracy on words with typical prosody (first-syllable stress) compared to with atypical prosody (second-syllable stress) (Dollaghan et al., 1993; Roy & Chiat, 2004).

Finally, in the view of NWR as an index of phonological STM (S. E. Gathercole & Baddeley, 1990), word length, as measured by the number of syllables in a word, is the most commonly reported predictor variable in NWR tasks is. This is because phonological STM load increases as the number of syllables increases (S. E. Gathercole & Baddeley, 1990; Szewczyk et al., 2018). In previous studies, as the number of syllables increased from two to five, NWR accuracy decreased in both monolingual and bilinguals (Boerma et al., 2017; Chiat, 2015; dos Santos & Ferré, 2018; Gathercole, Willis, Emslie, & Baddeley, 1992; Thordardottir & Brandeker, 2013).

In the present study, we examined how emergent bilingual children’s performance on the comprehension and production of single word vocabulary is influenced by word class
(nouns vs verbs) and subjective AoA, as estimated by native speakers of English. We also examined how performance on NWR tasks targeting quasi-universal and language specific properties is influenced by sublexical properties, such as Ngram frequency, transitional probability, and prosody, by phonological STM, measured by the number of syllables and consonants in the nonword, as well as by phonological or articulatory complexity, measured by presence of consonant clusters in nonwords.

**Relationship between vocabulary and NWR**

Whether the sublexical processing and memory skills implicated during a NWR task are driving forces behind vocabulary growth or the result of a larger vocabulary is highly debated in the literature (Metsala, 1999; Verhagen et al., 2019), and the direction of their relationship seems to change as a function of age (Rispens & Baker, 2012). NWR can be a predictor of L1 lexical acquisition in the early years (Gathercole et al., 1992). As children get older, however, the strength of the relationship between NWR and vocabulary development seems to decline. The strong relationship between NWR and vocabulary development seems to fade away after the first school year and other cognitive skills, with long-term memory (LTM) playing a more important role in later L1 vocabulary development. This is because older monolingual children can rely on lexical and sublexical knowledge available in their LTM. This knowledge supports then STM and facilitates vocabulary learning thanks to access to this lexical and sublexical information (Archibald & Gathercole, 2006; Metsala, 1999; Verhagen et al., 2019). Although a NWR task as a measure of phonological STM abilities does not rely on lexical materials and minimizes the influence of long-term lexical representations, it engages other sublexical representations. With more language exposure, familiarity with more frequent combinations of phonemes increases, so larger chunks of phonemes are learnt. This means that learning new words or nonwords is easier because the working memory is
maintaining fewer, longer phoneme chunks (Jones, 2016). In previous research, differences seen in NWR task performance between phoneme chunks with high and low probability were best predicted by vocabulary size, showing that vocabulary size may reflect lexical and sublexical knowledge (Munson et al., 2005; Szewczyk et al., 2018).

This relationship also may explain why phonological skills both improve and are improved by vocabulary size in bilingual children (Masoura & Gathercole, 1999; Paradis, 2011). Larger receptive vocabulary in both monolingual and bilingual children significantly predicted their ability to learn real and nonwords in the target language (both the L1 and the L2) (Kan & Kohnert, 2012; Roy & Chiat, 2004; Sorenson Duncan & Paradis, 2016; Szewczyk et al., 2018). Earlier and longer exposure to the language tested may explain this connection, as they improve both vocabulary and NWR (Antonijevic-Elliott et al., 2020; Gathercole et al., 2013; Sorenson Duncan & Paradis, 2016). In a study relevant for our purposes, Nicolay and Poncelet (2013) reported that NWR is differentially related to minority L2 expressive and receptive vocabulary in the first three years of French immersion primary education in Canada. Namely, the link between NWR and L2 expressive vocabulary was stronger and lasted longer than the link between NWR and L2 receptive vocabulary. Finally, bilingual children with the same L1 vocabulary sizes as monolingual children were significantly better at learning nonwords and their meanings than monolingual children, suggesting that there may be a bilingual effect for nonword learning. This may point to bilingual children being more efficient when encoding phonologically unfamiliar information (Kaushanskaya & Marian, 2009).

**Individual differences and development of vocabulary and NWR in bilingual children**

Vocabulary development is a language domain that correlates highly with the development of other language and cognitive domains and academic achievement. At the same time,
vocabulary knowledge, operationalised as receptive and expressive vocabulary size (that is, the number of words a speaker comprehends and produces) is highly susceptible to the amount and quality of input an individual receives in both monolingual (Weisleder & Fernald, 2013) and bilingual contexts (Chondrogianni, 2018 for a review). Studies with bilingual school-age children attending majority L2 programmes have shown that those with a later age of L2 exposure (AoE) have a steeper L2 vocabulary learning spurt compared to their bilingual peers with a lower L2 AoE (Chondrogianni & Marinis, 2011; Golberg, Paradis, & Crago, 2008). This has been attributed to two factors: (i) to children’s cognitive maturity when exposed to the L2 at a later age, which allows them to consolidate lexical information faster, and, (ii) to the positive, in this case, interdependence between languages that allows children to build on concepts already acquired in their first language (Golberg et al., 2008) (but also see (Chondrogianni, 2018; Unsworth, 2013) for a discussion of studies reporting negative interdependence across different language domains in simultaneous and sequential bilingual children).

The language of schooling has been shown to interact with the home language to affect vocabulary outcomes in both the majority and the minority language. Studies have shown that although the language of schooling affects vocabulary size, the language used at home will have larger effects on vocabulary development, especially for minority languages, as it reinforces the contexts outside the school context where the minority language can be used, (Dijkstra et al., 2016; V. C. Gathercole et al., 2008; V. C. Gathercole & Thomas, 2009; Hoff et al., 2014; O’Toole et al., 2019; Smithson et al., 2014). Additionally, speaking one of the two languages at home prior to attending immersion education may allow more time for the L1 vocabulary skills to be consolidated before the onset of L2 education. This is in line with studies reporting that children with later majority L2 immersion showed greater gains in the minority L1 (Schwartz et al., 2012).
Turning to studies investigating the majority language vocabulary skills in children attending minority language immersion education for revitalisation, two studies with Welsh-English bilingual children from two to fifteen years of age showed that older children with higher exposure to English at home had higher vocabulary scores on a standardised assessment of single word vocabulary comprehension for British English (V. C. Gathercole et al., 2013; Rhys & Thomas, 2013) compared to more Welsh-dominant peers in Welsh-immersion education, who reached age-appropriate English receptive vocabulary by adolescence. In the study Gathercole & Thomas (2009) with Welsh-English bilingual children using a non-standardised shorter version of the Welsh and English standardised receptive vocabulary tasks adopted in the studies mentioned above, the authors reported that seven-year-old bilingual children from English speaking homes attending Welsh-medium schools performed on a par with children with the same home language background attending Welsh-English bilingual schools. Note that at the age of seven years the majority language English is not supported in Welsh-medium education. These studies from the Welsh context suggest that exposure to the majority language outside the school context (in the home in this case) suffices to offset any lack of instruction in the school setting, even when standardised assessments are used to assess the majority language.

Whether performance on NWR is susceptible to exposure factors is more controversial, and results depend on the design of the NWR tasks and what exactly they measure (Chiat, 2015). Generally, bilingual children perform better on NWR tasks compared to vocabulary or morphosyntactic tasks, which tap into LS knowledge (Chiat, 2015; Thordardottir & Brandeker, 2013). Bilingual children’s performance on NWR also improves with age (Chiat & Polišenská, 2016b; Sorenson Duncan & Paradis, 2016), although longer LoE to the majority L2 is also an important predictor for L2 NWR performance, and in some studies, over and above the child’s current age (Sorenson Duncan & Paradis, 2016).
Numerous recent studies with bilingual children learning the majority L2 have shown that, when children are assessed on NWR tasks designed to not favour L2-specific sublexical properties, such as LS syllable structure and phonotactics, or to avoid language specific consonant clusters, such as the case of Quasi-Universal (QU) NWR task developed by (Chiat, 2015; Chiat & Polišenská, 2016b), bilingual children perform better on these tasks than on the LS ones (Boerma et al., 2015), and may not differ from their monolingual peers, especially after controlling for L2 receptive vocabulary (Antonijevic-Elliott et al., 2020; Chiat & Polišenská, 2016b; Engel de Abreu et al., 2012; Meir & Armon-Lotem, 2017). In two other studies using NWR tasks similar to the present study, current age was not a significant predictor for the QU and the LS NWR task performance in the bilingual and monolingual children (Antonijevic-Elliott et al., 2020; Szewczyk et al., 2018); conversely, earlier AoE to the majority L2 (English) in the case of bilingual children, and receptive vocabulary in the case of monolingual children, were significant individual variables in each study respectively, although the effect of age may have been masked by the stronger vocabulary differences between the groups (Szewczyk et al., 2018).

Finally, individual differences on vocabulary learning and NWR can be modulated by nonverbal intelligence, as a proxy for analytic reasoning. Nonverbal IQ has shown to significantly and positively correlate with receptive vocabulary in both the L1 and L2 (Daller & Ongun, 2017; Paradis, 2011), but has weak or no correlations with NWR tasks (Conti-Ramsden et al., 2001; Szewczyk et al., 2018). The discrepancy between two measures may be attributed to the close link between analytic reasoning and vocabulary skills reported for both monolingual and bilingual contexts (Paradis, 2011).

It is important to note that in previous studies investigating the development of nonword repetition in bilingual children of primary school age, these bilingual children were tested in the majority L2 which is supported at school, for example, on L2 English, and the
child’s L1 was different from the majority L2. These studies differ from our study where English is the children’s L1 but is not supported in the school context. In the present study, and because we are focusing on English as the majority language of the home and the wider community, we expect exposure variables (LoE and AoE) to play a lesser role for the children in our sample, especially in the case of the QU NWR task, whereas age at the time of testing will remain a contributing factor, as it also affects monolingual children’s performance on NWR tasks. We return to more detailed predictions in the relevant section below.

**Present study**

In the present study, we examined the individual, linguistic and item-related factors that affect the comprehension and production of English vocabulary and NWR in English-Gaelic bilingual children in GMPE. In GMPE, the first three years (P1-P3) of primary school (age five to eight years) are taught in L2 Gaelic, after which point L1 English is introduced into the curriculum (O’Hanlon et al., 2012). The two year-groups that we targeted in the current study attended primary years 2 and 3 at the time of testing, what is considered to be part of the first level in Scottish National Curriculum for Excellence (Scotland, 2021). Currently, how these factors affect majority language acquisition in GMPE remain to date unexplored. Specifically, we asked the following research questions:

1. How do item-level factors, such as word class and AoA, and child-level factors, such as current age, length (LoE) and frequency of exposure (FoE) to English and Gaelic currently and at an early age, as well as non-verbal IQ modulate English-Gaelic bilingual children lexical production and comprehension?
2. How do item-level (word length, Ngram frequency, prosody, and phonological complexity) and child-level factors affect English-Gaelic children’s ability to repeat nonwords that have language specific and quasi-universal properties?

3. What is the relationship between English vocabulary and NWR in English-Gaelic bilingual children attending the first level of GMPE?

Predictions

Starting with the item-level factors for vocabulary, we expected children to have better receptive than expressive skills, and to perform better on nouns and words with lower AoA compared to verbs (Altman et al., 2017; Haman et al., 2017). Better performance on comprehension than on production was expected because lexical retrieval and articulation implicated during production are more effortful than accessing lexical or phonological representations, implicated during comprehension (Gibson et al., 2014; Levelt, 2001). We also expected children to perform better on nouns than on verbs not only because (concrete) nouns are generally acquired earlier and are easier to depict than events (McDonough et al., 2011), but also because nouns tend to have less marked morphosyntactic and distributional properties (Haman et al., 2015; Haman, Łuniewska, et al., 2017), which may impact on acquisition. Finally, we expected words with lower AoA to elicit higher performance as children would have had more time to practice with earlier than with later acquired words (Łuniewska et al., 2016, 2019).

Turning to the NWR task, we expected that nonword length as a measure of phonological STM will influence performance across the different NWR subtasks. Specifically, we expected words with fewer syllables to have higher accuracy than words with more syllables if increase of word length negatively impacts on phonological STM (Chiat, 2015; S. E. Gathercole & Baddeley, 1990). In terms of the sublexical properties of
nonwords, we anticipated that participants would perform better on the QU with prosody (PS) set than the QU set without prosody given the facilitating role of prosody in (non-)word learning (Archibald & Gathercole, 2007; Archibald et al., 2009; Roy & Chiat, 2004). We also expected performance on the QU set to be better than that on the LS set due to the overall less marked and articulatory easier nonword structure in the QU set (CVCV) and the absence of consonant clusters compared to the LS set (Boerma et al., 2015). In the LS set, we expected Ngram frequency, as a subtle measure of sublexical representations, syllable number, as a measure of phonological STM, and receptive vocabulary scores, as a measure of the child’s English lexical knowledge, to be significant predictors. Specifically, we predicted that children would perform better on items that have a higher Ngram frequency, fewer consonants (Szewczyk et al., 2018), fewer syllables (Boerma et al., 2015), and typical prosody (Roy & Chiat, 2004). We also expected lexical and phonological skills to be positively associated, as the ability to repeat nonwords in younger children is related to their ability to learn new words (Masoura & Gathercole, 1999). In addition, learning new words (i.e., increase in vocabulary size) should provide more phonotactic/sub-lexical knowledge and make NWR easier (Szewczyk et al., 2018).

Turning to the child-level factors affecting performance on vocabulary, we predicted that older children with higher IQ (Daller & Ongun, 2017), earlier, longer, and more frequent exposure to the majority L1 English and later and less frequent exposure to the minority L2 Gaelic (Armon-Lotem & Ohana, 2017) would perform better on the English expressive and receptive vocabulary tasks (Gathercole et al., 2013). This is because the more exposure bilingual children have to consolidate the home language vocabulary (English in this case), the better they perform on related tasks (V. C. Gathercole et al., 2013; Golberg et al., 2008). However, we expected that the strong language exposure effects generally reported for other bilingual children, especially for bilingual children of a heritage background attending
majority schools (Chondrogianni, 2018), may not be as pronounced in the present study. Specifically, reduced language exposure may lead to weaker links between lexical access and phonological representations. These weaker links have been reported for bilingual children in previous studies when tested in their minority L1 or majority L2 (Gibson et al., 2020). However, these effects may not be observed to the same degree in the present study, where the emergent children were coming from predominantly majority L1 English-speaking homes and were tested in their majority home language that they were regularly using outside the school context.

For the NWR task, we expected that older children would perform better than younger children (Chiat & Polišenská, 2016), when they have higher receptive vocabulary scores (Szewczyk et al., 2018) and when they have higher exposure to the language tested (majority language in this case) (Antonijevic-Elliott et al., 2020; Sorenson Duncan & Paradis, 2016), as children’s STM skills and sublexical knowledge improves with age and language exposure. If bilingualism enhances STM skills, we anticipated children who have more bilingual exposure to perform better on the NWR task (Morales, Calvo, & Bialystok, 2013). We also expected that performance on the NWR task may also be driven by overlap in sublexical properties between Gaelic and English and specifically, prosody. Namely, given that Gaelic words also display a trochaic stress pattern (Lamb, 2001; Nance & Maolalaigh, 2019), we expected this overlap in language-specific sublexical properties between the two languages to reinforce performance on the nonword repetition task conditions that tap into prosody. We also expected language exposure to differentially affect correct repetitions across the three subtasks of the NWR task. This is because, the QU subtests were constructed to minimise similarity with any specific language and have a CV structure that is present in many different languages. As a result, they should be the least sensitive to language exposure. In contrast, we expected the language specific subtasks to be more susceptible to language
exposure.\(^1\) and that higher exposure in English (L1) would lead to better performance in both LS NWR task and the vocabulary task (Paradis, 2011; Szewczyk et al., 2018).

**Participants**

Fifty-five children (31 female) aged between six and eight years old were recruited from three Gaelic-Medium primary schools in three different locations in Scotland, UK, that is Edinburgh, Glasgow, and the Isle of Skye. Parents provided information about children’s language use and history using the Questionnaire for Parents of Bilingual Children (PABIQ) (Tuller, 2015). All children were exposed to the majority L1 English from birth and had varied exposure to the minority L2 Gaelic outside the school context. Twenty-four children were exposed to Gaelic before their third birthday and all children had large variation in their frequency of exposure to Gaelic (FoE-GA) at an early age (early FoE-GA) and before attending primary school, which may or may not have coincided with nursery attendance. In contrast, children’s exposure to English (FoE-EN) at home was relatively high, both at an early age and during primary school. As English was the dominant language for all children, we used the Clinical Evaluation of Language Fundamentals (CELF-5) (Semel & Wiig, 2017) to identify and exclude children with potential Developmental Language Disorder. Five children had a score of one standard deviation or greater below the age-appropriate norms, and, hence, were excluded from the final sample. Two other children were excluded because they did not complete all experimental tasks. No other children were excluded on account of reported hearing deficits or serious ear infections. This resulted in a final sample of 48 children. Descriptive data for the background variables can be found in Table 1.

\(^1\) We would like to thank one of the anonymous reviewers for pointing out this prediction to us.
Materials

Background assessments

Language and non-verbal abilities. To examine children’s general language abilities and to screen for language disorders, we administered the core component of the Clinical Evaluation for Language Fundamentals 5 (CELF-5 (Semel & Wiig, 2017). The core component of the CELF comprises the Sentence Structure, Word structure, Recalling Sentences and Formulating Sentences. Children’s raw scores were transformed into standard scores with a 90% CI. Children’s non-verbal IQ was assessed using the Raven’s Coloured Progressive matrices (Raven, 2003), which is suitable for children aged between five to eleven years. Children complete three sets of twelve puzzles, which required them to choose one of six options. The maximum raw score was 36. Raw scores were used for the statistical analysis in the present study to also account for the role of age.

Background questionnaire. To evaluate children’s exposure to English and Gaelic, we administered the Questionnaire for Parents of Bilingual Children (PABIQ) (Tuller, 2015). Parents reported at what age children were exposed to English and Gaelic and estimated how frequently (‘Always’, ‘Usually’, ‘Sometimes’, ‘Rarely’ or ‘Never’) English and Gaelic were spoken in the home before primary school attendance (early FoE-EN/GA) and currently (current FoE), which was converted into percentages (100%, 75%, 50%, 25%, 0%). In terms of early FoE-EN, only five reported that they usually (75%) spoke English at home, whereas the rest reported that they always spoke English at home. For early FoE-GA, the majority of parents reported that they sometimes (50%) spoke Gaelic in the home or usually (75%) speak Gaelic at home, whereas 15 reported that they rarely (25%, N=8) or never (N=7) spoke Gaelic at home. No parents/guardians in our sample were native speakers of Gaelic; they had mostly acquired Gaelic as an L2, if at all. Children had large variation in their AoE to Gaelic,
as well as how frequently they were exposed to Gaelic before formal schooling started and currently. Out of the forty-eight children in the sample, seven children had no early or current exposure of Gaelic in the home or early in the nursery, four children had exposure of Gaelic in the home but did not attend a bilingual English-Gaelic nurseries. All remaining children had some exposure to Gaelic in the home, albeit to variable degree, and attended bilingual English-Gaelic nurseries. Children’s current exposure to Gaelic and English (current FoE-EN/GA) was less variable, as most families spoke English at home, whereas input in Gaelic was provided in the school context (Table 1). We estimated children’s socioeconomic status (SES) based on the years of maternal education. Most mothers had a university degree or university and further training, and only four mothers had only secondary education or vocational training, suggesting that the children in the present sample belonged to mid-to-high SES (Table 1).

**Experimental tasks**

*Expressive and receptive vocabulary tasks.* To assess children production and comprehension of nouns and verbs we used the English version of the Litmus Cross-Linguistic Lexical Task (CLT) (Haman et al., 2015). The CLT consists of four subparts: Nouns receptive/Verbs receptive, Nouns expressive/Verbs expressive. In the two CLT receptive subtasks (CLT-R), children saw four pictures and pointed to the picture which corresponded to the word they heard. In the two CLT expressive subtasks (CLT-E), children saw one picture and were asked to name it. Each set contained 32 items controlled for imageability and AoA (mean 3.98; SD 1.08; 1.99 – 7.56), as reported in Haman et al. (2015).

Although this task was originally administered to younger monolingual British English-speaking children aged between 5;2 and 6;9 years (Haman, Łuniewska, et al., 2017),
we decided to use the task with the older children in our study for the following reasons. First, the adoption of the English CLT was part of a larger project investigating the language abilities of English-Gaelic-speaking children in GMPE. As part of this project, we developed a Scottish Gaelic version of the CLT that had the same design criteria as the English CLT. The robustness and comparability of the research design between the English and the Gaelic tasks allow us to investigate how similar factors affect vocabulary development in the two languages of the bilingual individual. The results of this study are currently being published elsewhere. Second, the aim of the present study was not to establish norms on vocabulary comprehension and production or to examine the validity of this task as a measure of proficiency in this group of children (see adaptations of the CLT by van Wonderen and Unsworth (2020) with 4- to 9-year-old monolingual Spanish and Dutch-speaking children with that aim); rather, the aim of the study was to investigate what child- and item-level variables affect emergent bilingual children’s performance on vocabulary and nonword repetition, and how these variables may differentially modulate performance on these two related skills.

*NWR task.* Children’s phonological STM and sublexical skills were assessed using the British English version of the Crosslinguistic Nonword Repetition Task (CL-NWRT) (Antonijevic-Elliott et al., 2020; Chiat, 2015; Chiat et al., 2020; Chiat & Polišenská, 2016b). In this task, children saw a broken necklace on the computer screen and were asked to assist with making three new necklaces, bead by bead, by repeating ‘magic’ words. For each pre-recorded

\[\text{\textsuperscript{2}One reviewer raises the question about the appropriateness of the task with the age group in the present study. Although we acknowledge the limitations of using this task with this group (see Discussion), the emergent bilingual children in our study are different from the children in the Haman et al. (2017) study in various ways. First, the children in our study were older than the ones in the Haman et al. (2017) study. Despite that, they did not have higher performance than the younger monolingual children described in (Haman, Luniewska, et al., 2017) study. Second, the children in our study are not monolingual; they are emergent bilingual children. And third, they are attending Gaelic-medium immersion education and not English mainstream education. Finally, the children in our study exhibited variability on subparts of the task that allowed us to address our research questions.}\]
‘magic’ word that children heard and repeated back, a bead appeared on the screen, regardless of the accuracy of the repetition. Children only heard each non-word once. The CL-NWRT comprises three sets: the Quasi-Universal (QU), the QU with Prosody Specific (PS), and the Language-Specific (LS) set. Each component had different items with item-specific variables such as number of syllables and prosody being manipulated (see Table 2 for more detail). In the LS task, Ngram frequency and transitional probability were also manipulated following Chiat (2015).

INSERT TABLE 2 ABOUT HERE

Both the Litmus CLT and the CL-NWRT were developed during the COST Action IS0804 “Language impairment in a multilingual society: Linguistic patterns and roadmap to assessment” (2008-2013, www.bi-sli.org).

Procedure

Children were tested individually in a quiet room in their primary school over three sessions on separate days by a Speech and Language Therapist and two research assistants. The background and experimental tasks were presented in a randomised order. In all experimental tasks, the words were pre-recorded by a British English female speaker. All children’s responses to the tasks were recorded on a digital voice recorder and later transcribed. This study received ethical permission from the Research Ethics Committee in the School of Philosophy, Psychology and Language Sciences at the University of Edinburgh, UK (21-1920/1).

Coding and Scoring
All responses to the CLT-E were coded as correct following the CLT instruction manual (Haman et al., 2015) (see supplementary material for examples). The CL-NWR task responses were transcribed into the International Phonetic Alphabet and scored by an independent rater. A second independent rater reviewed 10% of the transcriptions, chosen randomly. The interrater reliability whole-item rate was 95.5%. In the present study, we focused on whole-item scoring (Chiat, 2015). If the child’s response had all and only the same phonemes in the correct order as the target nonword, it was marked as correct e.g. dula - /'du,lɑ/. Any additions (/’du,lɑm/), omissions (/’u,lɑ/) and substitutions (/’du,mɑ/) were marked as incorrect (Antonijevic-Elliott et al., 2020; Szewczyk et al., 2018).

Data Analysis

Data was analysed using R-Studio (version 3.6.0) (R Core Team, 2021) and the R packages “lme4” (Bates et al., 2015), “Hmisc” [version 4.4-0 (Harrell & Dupont, 2020)] and “MuMIn,” [version 1.43.6]. For each research question, a generalised logistic mixed model was constructed. Because of the large number of predictors, testing the maximal model with all fixed effects and random effects was not possible. Therefore, we used an overfitting anti-conservative model which contained all variables as fixed effects and only by-participant and by-item random intercepts (Szewczyk et al., 2018). We decreased the number of variables using a backwards stepwise selection procedure. At each step, we compared alternate models with each variable removed in turn and chose the model with the lowest Akaike Information Criterion (AIC), which reflected a better goodness of fit while accounting for the complexity of the model. When the AIC value no longer decreased, the most important predictors were found, and this was the final model for the task. Interactions or non-linear contributions were found by building a further set of models and comparing to these final models through likelihood ratio comparisons. Marginal and conditional $R^2$ values of each final model were
calculated using the function “r.squaredGLMM” from the package “MuMIn”. The marginal R² indicated how much variance was explained by the fixed effects whilst the conditional R² indicated the variance explained by both fixed and random effects.

## Results

### Initial analysis: Collinearity checks

Prior to constructing the various models, we examined whether there was collinearity among the various predictor variables, which could cancel out potential predictor outcomes. For the item- and participant-related predictors (Table 3), we looked at the Spearman correlation coefficients as not all variables had normal distribution (Tables 4 and 5).³ There was high collinearity between certain participant-related predictors, for example, between AoE-Gaelic, LoE-Gaelic and early FoE-Gaelic. There was also a high correlation between certain item-related predictors, such as between Ngram Frequency and Number of Syllables. As each variable measured different observations and could uniquely contribute to the CLT and NWR task performance, we decorrelated any variables correlated with a rho-value ≥ 0.7. Where two variables (A and B) were highly correlated to each other and only to each other, variable A was converted into residuals of a linear model (A~B) and variable B was left intact. If variable B was taken out of the model, the residuals of variable A were then replaced by the original variable A. Where three variables (variable A, B and C) were highly correlated to each other, two models were made, one with A and B and one with A and C. One of each pair was converted into residuals as above, then final models were compared using AIC.

³ Age of Exposure to English, Length of Exposure to English and Level of Maternal Education as a proxy for socioeconomic background (SES) were included in the initial analysis. However, these factors did not correlate or predict any of the outcomes, and, hence, they were not included in the results reported here. Specific to SES, and given that previous studies have reported effects of SES on the tasks examined in the present study (e.g., (Luniewska et al., 2016; Meir & Armon-Lotem, 2017; Roy & Chiat, 2004), we believe that the lack of SES effect in the present study is due to the rather homogeneously high SES of the children in our sample and the lack of variation (Table 1; mean years of maternal education = 17.17; SD = 2.63). Even the four children with lower SES performed within the range of the entire group, and hence were included in the analysis.
Crosslinguistic lexical task

On average, children had 91% accuracy across the receptive and expressive CLT tasks (83-98%; SD 4%). Items were correctly answered by 91% of children (17-100%; SD 17%). This shows that most children answered most items correctly, which indicates a ceiling effect, especially when observing that children reached the maximum probability of answering correctly (1.00) and that there was a smaller variance in the receptive tasks (see Table 6 and Figure 1).

To understand the relative and unique contribution of the item- and child-related variables to the CLT-R and CLT-E separately and to avoid reporting here multiple three-way interactions between categorical (Task Type) and continuous background variables (e.g., AoE) that would need to be further unpacked, we constructed different models for each modality after checking for initial interactions. This decision was further reinforced by an original analysis that revealed an interaction between Word class and Task Type and a significantly higher performance on the CLT-R compared to CLT-E ($t(47) = 14.512, p < 0.001$).

The best models for predicting performance on the CLT-R and CLT-E tasks are presented in Table 7. Across all models, the item-related variables, such as Word Class and AoA, as well as the participant variance (conditional $R^2$), explained more of the variance than
the child-related variables, such as AoE, which in both models contributed to their explanatory variance but was not a significant predictor. Additionally, the model for CLT-E had greater explanatory adequacy than CLT-R due to children’s ceiling accuracy on CLT-R.

INSERT TABLE 7 ABOUT HERE

NWR task

On average, children overall accuracy across the NWRT was 63% (32-88%; SD 14%). Items were correctly answered by 63% of children (19-95%; SD 21% SD) showing that there were no floor or ceiling effects (see Table 8 and Figure 2).

INSERT TABLE 8 ABOUT HERE

INSERT FIGURE 2 ABOUT HERE

To understand the relative and unique contribution of variables in children’s performance on the NWR tasks, we compared performance across the three different NWR sets. Children scored significantly higher on the QU set than the PS or LS set (p < 0.001). Because performance differed so greatly between the sets, and the LS set had different item-related variables to QU and PS sets, three models were built, one each for the QU, PS and LS sets. Furthermore, to answer research question 3, namely whether receptive and expressive vocabulary scores contributed to the NWR performance, we entered the vocabulary scores (percentage accuracy per participant) as fixed effects in the models.

INSERT TABLE 9 ABOUT HERE
Table 9 presents the best fitting models for predicting performance on the NWR QU task, PS task and the LS task, respectively. Across all models, the item-related variables, such as number of syllables (QU and PS sets) and Ngram frequency (LS sets), as well as the inclusion of the participant variance (conditional $R^2$) explained more of the variance than child-related variables (LoE-GA and AoE-GA).

**Discussion**

In the present study, we examined the child- and item-level factors that affected the majority language receptive and expressive vocabulary and NWR skills of English-Gaelic six-to eight-year-old bilingual children attending Gaelic-medium immersion education. Specifically, we examined how emergent bilingual children’s performance on the comprehension and production of single word vocabulary was influenced by word class (nouns vs verbs) and subjective AoA, as estimated by native speakers of English. We also examined how performance on NWR tasks targeting quasi-universal and language specific properties is modulated by sublexical properties, such as Ngram frequency, transitional probability, and prosody, by phonological STM, measured by the number of syllables and consonants in the nonword, as well as by phonological or articulatory complexity, measured by presence of consonant clusters in nonwords. We examined how these item-level factors interact with the child’s age and degree and age of exposure to the two languages.

**Lexical skills in bilingual children in immersion education**

For vocabulary, we expected the children in our study to perform overall better on the receptive task compared to the expressive task given the different cognitive demands between production and comprehension and further phonetic and articulatory processes imposed during production (Gibson et al., 2018; Levelt, 2001; Yan & Nicoladis, 2009). Whereas
access of lexical and phonological representations may suffice during comprehension, this is not enough for successful performance during production (Gibson et al., 2020), giving, thus, rise to a discrepancy in performance between the two tasks. We also expected children to perform better on nouns than on verbs given the well documented differences between the two word classes in their developmental trajectory (Altman et al., 2017; Haman, Wodniecka, et al., 2017; Sheng & McGregor, 2010), because concrete nouns are inherently morphosyntactically simpler, and due to the way objects and actions are depicted in picture-based tasks. The single-picture format of the task may favour nouns, because they label one item rather than a relational concept, as in the case of verbs, which is what makes nouns easier for children to learn when younger (McDonough et al., 2011). In addition, the depiction of events in the case of verbs may leave more room for naming errors, and this may cause the larger gap between noun and verb performance in the expressive tasks (Altman et al., 2017).

The results from the L1 English – L2 Gaelic bilingual children in our study were mixed. Although nouns overall elicited higher accuracy than verbs, children exhibited a discrepancy between comprehension and production only for verbs but not for nouns. The difference between receptive and expressive vocabulary skills only for verbs but not for nouns in this group of emergent bilingual children tested in their majority L1 differentiates them from other groups of bilingual children tested in their minority L1 or the majority L2. In these studies, receptive-expressive differences are reported for both word classes, albeit to a different degree (Altman et al., 2017; Gibson et al., 2012; Haman, Łuniewska, et al., 2017; Hansen et al., 2017; Oller et al., 2007; van Wonderen & Unsworth, 2020). The lack of discrepancy between the two modalities for nouns but not for verbs may be related to the older age of the children in the present study compared to previous studies, which led to overall higher accuracy compared with previous studies. However, the fact that the emergent
bilingual children in our study continued to have lower expressive skills for verbs reinforces the argument that inherent morphosyntactic properties of verbs, their different acquisitional trajectories from nouns and methodological considerations related to how verbs are depicted or the responses that they elicit (Altman et al., 2017; Haman, Łuniewska, et al., 2017; McDonough et al., 2011) lead to reduce performance.

We also expected subjective AoA to influence children’s comprehension and production accuracy (Łuniewska et al., 2016, 2019). Subjective AoA indirectly measures how long a child has been using a word and how entrenched the use of a word might be. This is because the earlier a word is acquired the more experience the speaker has with using this word in various contexts and the easier it is to access it. This may also explain why subjective AoA strongly correlates with parental reports on young children’s lexical development (Łuniewska et al., 2016). In the present study, subjective AoA significantly predicted performance on both the expressive and the receptive tasks. This is in line with previous studies with bilingual children (Altman et al., 2017; van Wonderen & Unsworth, 2020) but with what was reported in Haman et al. (2017) study with monolingual British English children, where AoA only correlated with verb production, even though the children in that study were younger than the ones in our study. This may be due to the different profiles of the children in the two studies: emergent bilingual children in our study versus entirely monolingual in the Haman et al. (2017) study. Additionally, in our study, AoA was a better predictor than the child’s current age, as it explained age effects while accounting for differences in difficulty across items. Interestingly, in previous studies which found current age a significant predictor, subjective AoA was not used as predictor (Armon-Lotem & Ohana, 2017; Gathercole et al., 2014). Since both current age and AoA are potential indicators of how much experience a child has had with a word, including them both as
predictors in a single model may lead to AoA winning over current age. Future studies would benefit from trying to disentangle the distinct contribution of these factors to word learning.

Turning to the child-level predictors, AoE to Gaelic, the minority L2, was a contributing, yet not significant, predictor in both receptive and expressive tasks. This may indicate that with later L2 exposure, the majority L1 English was spoken as the only language at home over a longer time, and hence, children had more time to consolidate their L1 vocabulary skills before being immersed in the minority L2, as reported in previous studies on bilingual children (Chondrogianni, 2018; Golberg et al., 2008). Frequency of exposure in the home language, which was also the majority language of the community, namely English, did not predict performance on majority language vocabulary. The relationship between home language use and vocabulary development is widely reported in studies where English is the majority L2 and there is more variation in its use at home because another minority language is spoken in this context (Armon-Lotem & Ohana, 2017; Gathercole et al., 2014). We were thus expecting exposure to modulate performance here as well, albeit to a lesser extent given the majority language status of the children in our sample. In the present study, however, and in contrast to previous studies, the use of English at home was quite high with little variation. Because the children in the present study displayed high performance on English vocabulary tasks and exposure to English was high as well, there was little variation in the data that is usually needed to give rise to correlations between two variables.

Given that in the present study, language exposure measures contributed to but did not strongly predict children’s accuracy on the expressive and receptive vocabulary tasks, the relationship between bilingual exposure and how it affects the vocabulary skills of this emergent bilingual group is harder to gauge. It seems that for the emergent bilingual children in our study the link between lexical access and phonological representations in the majority language remains strong, especially for nouns, and despite the lack of formal instruction of
the majority language in the school setting. This finding differentiates them from other studies with sequential bilingual children tested in their minority L1 or majority L2 where weak links between lexical access and phonological representations have been reported, giving rise to receptive-expressive differences (Gibson et al., 2020; Gollan et al., 2008). At the same time, the six-to-eight-year-old bilingual children in our study exhibited similar performance to the monolingual children in the Haman et al. (2017) study, especially in the production task where their performance was not at ceiling (approx. 76% on verbs). This is despite the fact that the children in our study were two years older than the ones in the Haman et al. (2017) study. Future studies could investigate whether and under which conditions the link between lexical access and phonological representations becomes less strong in the majority language, if at all, for bilingual children in immersion education, and whether certain word classes are affected more than others.

Finally, nonverbal IQ was a significant predictor only of expressive vocabulary skills, with children with higher Raven’s scores having larger expressive vocabularies. The relationship between non-verbal IQ and vocabulary development is well documented in monolingual studies (Lervåg et al., 2019) and has been confirmed in previous studies in bilingual children (Daller & Ongun, 2017). In this study, the relationship between nonverbal IQ and expressive vocabulary may be due to the higher variance in performance in the latter compared to the receptive task. Namely, and as displayed in Figure 1, the receptive vocabulary component elicited almost ceiling performance with little variation, whereas the expressive vocabulary component of the task gave rise to more variation. Given that there was also more variation in the nonverbal IQ task, as displayed in Table 1, the expressive part of the CLT and the nonverbal IQ task correlated more than the receptive CLT component and the nonverbal IQ task in the present study.
Overall, in this study, item-related variables significantly modified the performance of English-Gaelic emergent bilingual children, as the coefficients in the models showed, whereas participant-related variables were important for improving the fit of the model but were not significant themselves. This may be due to the fact that children’s vocabulary skills in the majority language were already high and, with the rather homogeneous degree of exposure to the majority language in the home, the current variation in the present sample gave rise to exposure as a contributing but not a significantly modulating factor.

**NWR skills of bilingual children in immersion education**

In the NWR task, we predicted that children’s accuracy will be affected by the number of syllables in a nonword, a measure of phonological STM (S. E. Gathercole & Baddeley, 1990). These predictions were borne out for the QU and PS sets: as the number of syllables increased, the phonological STM load increased, leading to lower accuracy on lengthier nonwords (Boerma et al., 2015; dos Santos & Ferré, 2018; S. E. Gathercole et al., 1992; Szewczyk et al., 2018; Thordardottir & Brandeker, 2013). However, this was not the case for the LS set, where number of syllables was not a predictor. This may be related to the interplay between NWR as a measure of phonological STM and the sublexical properties that we examined in the present NWR task, and which we discuss in more detail below.

Looking at the sublexical properties more closely, we found that simple CVCV structure and lack of consonant clusters in the QU set led better performance compared with the PS and the LS sets (Boerma et al., 2015; Chiat, 2015). However, our results did not confirm our prediction that children would be facilitated by prosody, and hence, perform better on the PS than the QU set. With very early and high exposure to the majority L1 English at home and in the community, we expected that children in this age range would perform better on nonwords with typical prosody (PS) than nonwords with even stress (QU).
given the facilitatory role of these sublexical properties and as it is also reported in previous studies with monolingual (Archibald et al., 2009; Archibald & Gathercole, 2007; Roy & Chiat, 2004) and bilingual children (Boerma et al., 2015). The significantly better performance on the QU set than the PS set in the present study could be because the administration of the three sets was fixed and children always completed the QU task after the PS task. Because the two tasks contain the same nonwords, repetition possibly improved words in the second set (QU) despite the lack of prosody.

Typical prosody was also not a predictor in the final model for the LS set. In previous studies, children performed better on nonwords with typical prosody compared to atypical prosody (Dollaghan et al., 1993; Roy & Chiat, 2004). This different result might be because Ngram frequency could have explained more variance in this study, as it varied by item, providing a subtler measure of sublexical properties, whereas prosody varied only across sets.

Finally, we did not see consonant clusters modulate performance on the LS set, contrary to what was reported in previous studies with monolingual children (Szewczyk et al., 2018). This could be because the consonant clusters in the present English LS task were not as challenging as in other tasks reported in previous studies (Gathercole et al., 2012; Szewczyk et al., 2018). Furthermore, the medium-to-high correlation between number of syllables, consonants and Ngram frequency in the current study suggests that Ngram frequency may explain variance that number of syllables or consonants may also explain (Gathercole et al., 2012). This may be because as we learn larger chunks of phonemes that occur more often in the language (Jones, 2016), working memory is able to more easily maintain fewer, longer and familiar phoneme chunks found in higher Ngram non-words, than many, short and less familiar phoneme chunks (from non-words with low Ngram frequency) (Szewczyk et al., 2018). Such an explanation would be compatible with phonological models, such as the EPAM-VOC and CLASSIC, that assume that the presentation of nonwords
activates sequences of phonemes in the long-term memory that are present in the nonword (e.g., see (Jones et al., 2007).

Turning to the child-level predictors, we predicted that increased age, receptive vocabulary scores and exposure to the majority language would predict better LS set performance but may not affect performance on the QU and PS sets, as these two tap into more universal phonological properties of (non)word structure (Chiat & Polisenksa 2016; Szewczyk et al., 2018). Looking at each individual variable in turn, current age was not a predictor in any final model of the NWR task. This was in line with our prediction, as we were not expecting any child-level variables to be significant for the QU set. In the PS and LS sets, we anticipated current age to improve performance. However, it seems that, in this study, other factors such as AoE to L2 Gaelic and receptive vocabulary scores were more significant child-related predictors, similarly to previous studies (Antonijevic-Elliott et al., 2020; Szewczyk et al., 2018). Interestingly, LoE to the minority L2 Gaelic and AoE to L2 Gaelic were significant predictors of children’s performance on the NWR task, albeit in surprising ways. That is, earlier and longer exposure to minority L2 Gaelic rather than to the majority English significantly predicted better performance on the PS and LS sets. Although one may assume that more exposure to the majority language and less exposure to minority L2 would lead to better performance in the PS and LS sets, as they contained prosody and/or phonetics typical of the majority L1, we should note that Gaelic is also a stress-timed language with a trochaic pattern similar to English (Morrison, 2019) and with more multisyllabic words than English (Lamb, 2002). The same children tested in this study were also administered the Gaelic version of the CLT, where we established that they know a range of Gaelic words (including multisyllabic ones) in Gaelic (approx. 76% for comprehension and 40% for production). This ability to comprehend and produce multisyllabic words in another language may have reinforced their understanding of the
prosodic patterns in the two languages, as well as their performance on multisyllabic words in the NWR task. This may have resulted in improved performance on these prosody-specific tasks with increased and earlier exposure to this particular minority language (Kaushanskaya & Marian, 2009). Additional evidence for this also comes from how exposure affected performance on the QU set. Children’s performance on the QU set was also influenced by exposure variables in that children with earlier exposure to the minority L2 Gaelic and less current exposure to the majority L1 English performed better on the QU set. Although we did not expect exposure to affect this component of the CL-NWRT (Antonijevic-Elliott et al., 2020; Chiat & Polišenská, 2016b; Engel de Abreu et al., 2012; Meir & Armon-Lotem, 2017; Sorenson Duncan & Paradis, 2016), it seems that, in our study, earlier bilingual exposure led to better performance on a non-language-specific task.

Finally, we found that expressive but not receptive vocabulary skills contributed to children’s performance on the LS set of the NWR task, which carries language-specific phonotactics. The lack of contribution of the receptive task to children’s performance on the NWR task reported in other studies (Sorenson Duncan & Paradis, 2016) may be due to children’s ceiling performance on this component, leading to little variation, and, hence, lack of a significant effect. Given the greater variation observed in the expressive vocabulary task, this measure contributed more to children’s performance on the NWR task. It may also be because both tasks are expressive and hence involve similar processes of lexical or sublexical access and phonological representations, as well as articulation and oromotor movement, whereas the latter two are not implicated in comprehension (S. E. Gathercole et al., 1992; Levelt, 2001). Our findings are in line with previous studies that show larger vocabulary sizes led to better NWR skills, especially for children of this age (Kan & Kohnert, 2012; Roy & Chiat, 2004; Sorenson Duncan & Paradis, 2016; Szewczyk et al., 2018), although NWR
merely contributed to the model rather than strongly predicted lexical skills in our study (Verhagen et al., 2019).

**Conclusions**

In this study, we investigated how psycholinguistic, item-level and child-level variables affected performance of emergent bilingual English-Gaelic children attending Gaelic-medium education on vocabulary tasks targeting production and comprehension of single word vocabulary and on NWR tasks targeting quasi-universal and language specific properties. We found that children’s performance on both receptive and expressive vocabulary tasks was significantly better on nouns compared to verbs and on words acquired at a younger age. Children’s performance was significantly better on nonwords with fewer syllables and with higher Ngram frequency. Later exposure to Gaelic and higher nonverbal IQ contributed to better performance on the vocabulary task. Earlier exposure to Gaelic, less English spoken at home and higher expressive vocabulary size predicted better performance in the NWR task.

Taken together, these results reinforce the assertion that the extended exposure that majority language pupils have to the majority language (English) in an English-dominant community limits potential modulating effects of exposure to a minority language through immersion education. In the present study, this is evidenced in the context of a not previously examined minority language, Gaelic, undergoing revitalisation through immersion education. What is more, increased exposure to the minority language at home may lead to benefits in more general phonological skills related to word knowledge and word learning, such as nonword repetition, especially when sublexical properties of the two languages overlap.

**References**


https://doi.org/http://dx.doi.org/10.18637/jss.v067.i01


https://doi.org/doi:10.21832/9781783093137-008


https://doi.org/10.1044/2016_JSLHR-L-15-0293


Serratrice (Eds.), *Studies in Bilingualism* (Vol. 54, pp. 103–126). John Benjamins Pub Co. https://doi.org/10.1075/sibil.54.06cho


https://doi.org/10.1080/10489223.2016.1243692


https://doi.org/10.1177/0956797612443836


https://doi.org/10.1080/13670050.2020.1721427


https://doi.org/https://doi.org/10.1016/j.jml.2007.07.001


https://doi.org/10.1080/02699206.2017.1308553


Bilingual Children? Evidence from Polish-English Migrants to the United Kingdom.

*Frontiers in Psychology, 8*, 1444.


https://doi.org/10.1080/02699206.2017.1307455


https://doi.org/https://doi.org/10.1016/j.cognition.2016.04.017


https://doi.org/http://doi.org/10.1353/lan.0.0081

https://doi.org/10.1080/002075999399738

https://doi.org/https://doi.org/10.1111/j.1467-7687.2010.00968.x


https://doi.org/10.1037/0022-0663.91.1.3

https://doi.org/10.1080/13670050.2013.833892


https://doi.org/10.1037/0012-1649.39.4.706


Table 1: Participant background information

<table>
<thead>
<tr>
<th>Background variables</th>
<th>Mean (SD)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current age (months)</td>
<td>88.71 (6.87)</td>
<td>73 - 98</td>
</tr>
<tr>
<td>AoE-EN (months)</td>
<td>0.18 (1.03)</td>
<td>0 - 6</td>
</tr>
<tr>
<td>AoE-GA (months)</td>
<td>26.38 (20.13)</td>
<td>0 - 60</td>
</tr>
<tr>
<td>LoE-EN (months)</td>
<td>88.21 (6.79)</td>
<td>75 - 98</td>
</tr>
<tr>
<td>LoE-GA (months)</td>
<td>62.00 (21.37)</td>
<td>20 - 97</td>
</tr>
<tr>
<td>Early FoE-EN (%)</td>
<td>96.79 (8.36)</td>
<td>75-100</td>
</tr>
<tr>
<td>Early FoE-GA (%)</td>
<td>35.13 (22.55)</td>
<td>0-75</td>
</tr>
<tr>
<td>Current FoE-EN (%)</td>
<td>95.19 (8.31)</td>
<td>62.5-100</td>
</tr>
<tr>
<td>Current FoE-GA (%)</td>
<td>21.05 (16.25)</td>
<td>0-75</td>
</tr>
<tr>
<td>CELF-5 (standard scores)</td>
<td>105.21 (13.77)</td>
<td>86-147</td>
</tr>
<tr>
<td>Raven’s CPM Raw Score – (maximum score of 36)</td>
<td>23.40 (4.62)</td>
<td>15 - 33</td>
</tr>
<tr>
<td>Raven’s CPM percentile score</td>
<td>69.7 (24.64)</td>
<td>25 - 95</td>
</tr>
<tr>
<td>Maternal Education (years)</td>
<td>17.17 (2.63)</td>
<td>12 - 24</td>
</tr>
</tbody>
</table>

*Note. AoE = Age of exposure; LoE = Length of Exposure; FoE = frequency of exposure; EN = English; GA = Gaelic; CELF = Clinical Evaluation of Language Fundamentals; Raven’s CPM = Raven’s Colour Progressive Matrices.*
Table 2: Overview of the Crosslinguistic nonword repetition task (CL-NWRT) (British-English Version) (Chiat, 2015)

<table>
<thead>
<tr>
<th>CL-NWRT</th>
<th>Description of task</th>
<th></th>
</tr>
</thead>
</table>
| **Quasi-Universal (QU) set** | • 16 items equally divided across two, three, four and five syllables  
  • The QU set is designed to incorporate as many mutual features from different languages as possible  
  • Nonwords consist of syllables with simple consonant-vowel (CV) structure and the range of consonants/p, b, t, d, k, g, s, z, l, m, n/and vowels/a, i, u/that are common to many languages.  
  • Nonwords are produced with even length and pitch with the exception of the final syllable that is characterised by greater length and falling pitch to indicate the end of utterance.                                                                                                                                                                                                                                                                                                                                                                    |   |
| **Prosodically Specific (PS) set** | • 16 items equally divided across two, three, four and five syllables  
  • Includes the same items as the QU task but are given British English prosody (trochaic, Strong-Weak prosody)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |   |
| **Language-Specific (LS) set** | • 24 items equally divided between two, three and four syllables (eight items per syllable length)  
  • Phonotactic characteristics and prosody are typical of British English  
  • For each syllable length, items were divided equally into high and low phonotactic probability (four items per subtask)  
  • One of the four items in each of the six subtasks had one of the following properties:  
    o Atypical prosody (second-syllable stress) with no cluster  
    o Typical prosody (first-syllable stress) with no cluster  
    o Typical prosody with initial cluster  
    o Typical prosody with medial cluster                                                                                                                                                                                                                                                                                                                                                                                                                                                     |   |
Table 3: Participant- and item-related variables for the Crosslinguistic Lexical Task (CLT) and the Crosslinguistic nonword repetition task (CL-NWRT)

<table>
<thead>
<tr>
<th>Child-related variables for CLT and CL-NWRT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Child-Internal</strong></td>
<td></td>
</tr>
<tr>
<td>Current age (months)</td>
<td></td>
</tr>
<tr>
<td>Age of Exposure to Gaelic – L2 (AoE-Gaelic) (months)</td>
<td></td>
</tr>
<tr>
<td>Length of Exposure to Gaelic – L2 (LoE-Gaelic) (months)</td>
<td></td>
</tr>
<tr>
<td>Raven’s IQ test (raw score)</td>
<td></td>
</tr>
<tr>
<td>Mean CLT Receptive score (Ch-NWRT task only)</td>
<td></td>
</tr>
<tr>
<td>Mean CLT Expressive score (Ch-NWRT task only)</td>
<td></td>
</tr>
<tr>
<td><strong>Child-External:</strong></td>
<td></td>
</tr>
<tr>
<td>Early and current exposure to English at home (early/current FoE-EN)</td>
<td></td>
</tr>
<tr>
<td>Early and current exposure to Gaelic at home (early/current FoE-GA)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item-related variables for CLT and CL-NWRT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CLT (Receptive and Expressive subtasks):</strong></td>
<td>Word Class</td>
</tr>
<tr>
<td>Age of Acquisition of word (AoA) (years)</td>
<td></td>
</tr>
<tr>
<td><strong>CL-NWRT (QU and PS sets):</strong></td>
<td>Number of syllables</td>
</tr>
<tr>
<td><strong>CL-NWRT (LS set):</strong></td>
<td>Indices of length:</td>
</tr>
<tr>
<td>- Number of syllables</td>
<td></td>
</tr>
<tr>
<td>- Number of consonants</td>
<td></td>
</tr>
<tr>
<td>Indices of phonotactic probability:</td>
<td></td>
</tr>
<tr>
<td>- Ngram frequency</td>
<td></td>
</tr>
<tr>
<td>- Transitional probability</td>
<td></td>
</tr>
<tr>
<td>Typical/atypical prosody</td>
<td></td>
</tr>
<tr>
<td>Index of phonological complexity:</td>
<td></td>
</tr>
<tr>
<td>- Presence of consonant cluster</td>
<td></td>
</tr>
<tr>
<td>Variables</td>
<td>Age (months)</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>AoE English</td>
<td>-0.160</td>
</tr>
<tr>
<td>AoE Gaelic</td>
<td>-0.033</td>
</tr>
<tr>
<td>LoE English</td>
<td>0.994</td>
</tr>
<tr>
<td>LoE Gaelic</td>
<td>0.308</td>
</tr>
<tr>
<td>early FoE English</td>
<td>0.008</td>
</tr>
<tr>
<td>early FoE Gaelic</td>
<td>0.158</td>
</tr>
<tr>
<td>Overall NWRT</td>
<td>0.044</td>
</tr>
<tr>
<td>PS set</td>
<td>0.141</td>
</tr>
<tr>
<td>QU set</td>
<td>0.021</td>
</tr>
<tr>
<td>LS set</td>
<td>0.109</td>
</tr>
<tr>
<td>IQ (RS)</td>
<td>0.260</td>
</tr>
<tr>
<td>Overall CLT</td>
<td>0.195</td>
</tr>
<tr>
<td>CLT-R</td>
<td>0.091</td>
</tr>
<tr>
<td>CLT-E</td>
<td>0.217</td>
</tr>
</tbody>
</table>

Table 4. Spearman Correlation Coefficients for participant-related predictors in the Crosslinguistic Lexical Task (CLT) and the Crosslinguistic Nonword Repetition task (CLT-NWRT).
Note. AoE = Age of exposure; LoE = Length of exposure; FoE = Frequency of exposure; PS = Prosody specific; QU = Quasi-universal; LS = Language specific; RS = raw scores; CLT-R/E = Crosslinguistic Lexical task Receptive/Expressive
Table 5. Spearman Correlation Coefficients for item-related predictors in the Crosslinguistic Nonword Repetition task (CL-NWRT)

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Syllables</th>
<th>Consonants</th>
<th>Ngram Frequency</th>
<th>Transitional probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonants</td>
<td></td>
<td></td>
<td>0.818</td>
<td></td>
</tr>
<tr>
<td>Ngram Frequency</td>
<td>-0.719</td>
<td>-0.538</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitional prob.</td>
<td>-0.066</td>
<td>-0.136</td>
<td>0.600</td>
<td></td>
</tr>
<tr>
<td>Typical prosody</td>
<td>-0.054</td>
<td>0.153</td>
<td>0.049</td>
<td>0.202</td>
</tr>
<tr>
<td>Cluster</td>
<td>-0.134</td>
<td>0.369</td>
<td>-0.189</td>
<td>-0.141</td>
</tr>
</tbody>
</table>
Table 6. Mean (in %) accuracy in the Crosslinguistic Lexical Task (CLT)

<table>
<thead>
<tr>
<th>Task</th>
<th>Mean (SD)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLT overall (%)</td>
<td>0.91 (0.04)</td>
<td>0.83 – 0.98</td>
</tr>
<tr>
<td>CLT-Receptive</td>
<td>0.98 (0.02)</td>
<td>0.92 – 1.00</td>
</tr>
<tr>
<td>Nouns</td>
<td>0.99 (0.02)</td>
<td>0.94 – 1.00</td>
</tr>
<tr>
<td>Verbs</td>
<td>0.96 (0.03)</td>
<td>0.88 – 1.00</td>
</tr>
<tr>
<td>CLT-Expressive</td>
<td>0.85 (0.06)</td>
<td>0.68 – 0.97</td>
</tr>
<tr>
<td>Nouns</td>
<td>0.94 (0.05)</td>
<td>0.81 – 1.00</td>
</tr>
<tr>
<td>Verbs</td>
<td>0.75 (0.09)</td>
<td>0.55 – 0.94</td>
</tr>
</tbody>
</table>
Table 7. Best fitting models for the Crosslinguistic Lexical Task (CLT)

<table>
<thead>
<tr>
<th>CLT-Receptive</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>8.752</td>
<td>1.093</td>
<td>8.006</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Word Class</td>
<td>-0.816</td>
<td>0.548</td>
<td>-1.489</td>
<td>0.137</td>
</tr>
<tr>
<td>AoA</td>
<td>-0.817</td>
<td>0.219</td>
<td>-3.724</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AoE-GA</td>
<td>0.004</td>
<td>0.010</td>
<td>0.416</td>
<td>0.678</td>
</tr>
</tbody>
</table>

(Marginal $R^2 = 0.019$, Conditional $R^2 = 0.048$)

<table>
<thead>
<tr>
<th>CLT-Expressive</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.707</td>
<td>1.131</td>
<td>5.046</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Word Class</td>
<td>-1.614</td>
<td>0.458</td>
<td>-3.521</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AoA</td>
<td>-0.868</td>
<td>0.210</td>
<td>-4.139</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AoE-GA</td>
<td>0.005</td>
<td>0.006</td>
<td>0.732</td>
<td>0.464</td>
</tr>
<tr>
<td>Raven’s RS</td>
<td>0.055</td>
<td>0.029</td>
<td>1.891</td>
<td>0.059</td>
</tr>
</tbody>
</table>

(Marginal $R^2 = 0.173$, Conditional $R^2 = 0.378$)

Note. AoA = Age of acquisition word; AoE-GA = Age of exposure to Gaelic; RS = raw score.
Table 8. Descriptive statistics for the Crosslinguistic Nonword Repetition task (CL-NWRT)

<table>
<thead>
<tr>
<th>Task</th>
<th>Mean (SD)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL-NWRT overall (%)</td>
<td>0.63 (0.14)</td>
<td>0.32 – 0.88</td>
</tr>
<tr>
<td>Quasi-Universal (QU) set</td>
<td>0.70 (0.14)</td>
<td>0.31 – 0.94</td>
</tr>
<tr>
<td>Prosody-Specific (PS) set</td>
<td>0.61 (0.16)</td>
<td>0.25 – 0.94</td>
</tr>
<tr>
<td>Language-Specific (LS) set</td>
<td>0.60 (0.18)</td>
<td>0.21 – 0.92</td>
</tr>
</tbody>
</table>
Table 9. Best fitting models for the Quasi-Universal, the Prosody Specific and the Language Specific sets of the CL-NWR task.

<table>
<thead>
<tr>
<th>Fixed effect</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quasi-Universal set</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>7.370</td>
<td>1.716</td>
<td>4.295</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Number of Syllables</td>
<td>-0.838</td>
<td>0.191</td>
<td>-4.387</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AoE-GA</td>
<td>-0.005</td>
<td>0.007</td>
<td>-0.768</td>
<td>0.443</td>
</tr>
<tr>
<td>Early FoE-EN</td>
<td>-0.034</td>
<td>0.016</td>
<td>-2.137</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Marginal $R^2 = 0.155$, Conditional $R^2 = 0.262$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Prosody Specific set    |          |     |        |         |
| Intercept               | 1.842    | 0.914 | 2.016  | <0.05   |
| Number of syllables     | -0.645   | 0.216 | -2.983 | <0.01   |
| LoE-GA                  | 0.016    | 0.007 | 2.209  | <0.05   |
|                         |          |     |        |         |
| (Marginal $R^2 = 0.105$, Conditional $R^2 = 0.304$) |

| Language Specific set   |          |     |        |         |
| Intercept               | -4.890   | 2.370 | -2.063 | <0.05   |
| Ngram Frequency         | 0.747    | 0.264 | 2.833  | <0.01   |
| AoE-GA                  | -0.020   | 0.008 | -2.456 | <0.05   |
| CLT-E score             | 4.697    | 2.659 | 1.766  | 0.077   |
|                         |          |     |        |         |
| (Marginal $R^2 = 0.087$, Conditional $R^2 = 0.309$) |

**Note.** LoE-GA = Length of exposure to Gaelic; AoE-GA = Age of exposure to Gaelic; early FoE-EN = early frequency of exposure to English; CLT-E = Crosslinguistic Lexical Task - Expressive
Figure 1: Mean proportion accuracy on the CLT-Receptive/Expressive and overall
Figure 2: Mean proportion accuracy on the CL-NWRT (QU = Quasi-universal, PS = prosody-specific, LS = language-specific, and overall)