Methods in measuring vocabulary in UK schoolchildren:

is there evidence of a vocabulary gap?

Theodore Mills

Submitted to Swansea University in fulfilment of the requirements for

the Degree of Doctor of Philosophy in Applied Linguistics

Swansea University

2024

Abstract

There is a long-standing belief amongst academics, educators, and policymakers that there is a significant gap between the number of words that more privileged children know and the number that less privileged children know. It is believed that this 'word gap' or 'vocabulary gap' affects educational outcomes as well as children's overall wellbeing (Hart & Risley, 1995; Hoff, 2003, 2006; Oxford University Press, 2018; Quigley, 2018). If we wish to investigate this gap, then explicit descriptions of vocabulary development are important, both as a means of clarifying what a vocabulary gap might actually entail and for ensuring the vocabulary gap is effectively targeted (Durrant & Brenchley, 2019a).

This thesis presents a triangulation of different methods which can be used to quantify vocabulary development in children: (a) vocabulary size data collected from a cohort of children in Wales, (b) quantitative corpus analysis of a large-scale naturally occurring corpus of children's writing (the Oxford Children's Corpus), and (c) qualitative semi-structured interviews with schoolteachers. The results exemplify the difficulties involved in measuring vocabulary development in school-age children and have implications for curriculum development in the United Kingdom. In the discussion, these disparate strands of work are brought together using a language policy framework. The results are contextualised through a discussion of the vocabulary gap and how it has been imported into the United Kingdom from the United States. Particular attention is paid to critiques (Baugh, 2017; Burnett et al., 2020; Cushing, 2023; Kuchirko, 2019; Sperry et al., 2019b) of the original Hart and Risley (1995) study which gave rise to the idea of the word gap. Recommendations and implications for policymakers and academics are then presented.

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ....

Date .....14 August 2024.....

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ....

Date .....14 August 2024.....

List of Figures and Tables

Acknowledgements

I wouldn't have been able to complete this thesis without the unwavering support of my supervisor, Dr Vivienne Rogers, and my second supervisor, Dr Cornelia Tschichold, and the wider department of Applied Linguistics at Swansea. There were several moments I thought I wouldn't end up submitting anything at all but I'm pleased to say that I've finished a thesis I can be proud of. Big thanks to all participants at Grange Primary School, teachers and students, as well as to the rest of the teachers who contributed their valuable time to my research. Thank you as well to Oxford University Press for allowing me access to their corpus. The next acknowledgment goes to all my friends and family for their support. Especially Cadi and Josh—you've put up with my ranting about this PhD for years and for that I owe you. I truly am a lucky guy to be loved by you both. Another thank you to Clara, Magic, and Shadow, who perhaps have been my most constant sources of companionship over the last few years. You all supervised my work to perfection. Most of all I'd like to thank Chloe. Though you're no longer with us, you carried me here. Together we fought, against all odds, for the strength and courage to do all the things we ever wanted to do. This thesis wouldn't be here without you, and neither would I. Thank you.

What's written here and now is to be read there and then: some other time and some other place.

(Britton, 1972)

1. Introduction

The introduction to the national curriculum for English in England (Department for Education, 2014) states that English teaching is fundamental to the curriculum because it is the gateway to all other subjects. In the conclusions of their research review on the English curriculum, the Department for Education (2022) write:

> Vocabulary is fundamental to pupils' progress. Vocabulary is the foundational knowledge for reading, speech and writing. Narrowing the word gap between pupils who are word-rich and word-poor is vital. Vocabulary development and wider language development feed into reading and writing.

Vocabulary is linked to the key skills of reading and writing, as the report continues describing how "comprehension becomes possible partly through acquiring and deepening vocabulary" (Department for Education, 2022). These excerpts are examples of how vocabulary knowledge is emphasised in Anglophone school curricula. This emphasis is not unfounded: vocabulary affects what a person can do with a language and is important for many aspects of psycholinguistic processing. There is evidence in both first and second language acquisition research which shows that vocabulary is one of the key predictors of academic success and is linked to later literacy outcomes (R. C. Anderson & Freebody, 1981; Biemiller, 2011; Daller & Phelan, 2013; F. J. Duff et al., 2015; Lee, 2011; Pennebaker et al., 2014). Therefore, vocabulary deficits may limit a child's educational development (Biemiller, 2011; Bleses et al., 2016; Milton & Treffers-Daller, 2013). Furthermore, "it is a well-established finding in research on the reading comprehension of native speakers that vocabulary is the most important contributing factor [to improved reading comprehension]" (Read, 2000, p. 190). Vocabulary is also linked to phonological awareness and listening comprehension (Sénéchal et al., 2006).

The acknowledged importance of vocabulary has led to an increased number of quantitative studies on vocabulary development in the last decade (Durrant et al., 2021). Thus, research on vocabulary development has clear practical importance (Durrant & Brenchley, 2019b). However, our understanding of vocabulary development is constrained by a lack of research into vocabulary size, growth, and development in the early years as well as a number of methodological issues that must be addressed in quantitative vocabulary research. The lack of research is due to the challenge inherent in analysing first language vocabulary development, which is a multidimensional process and cannot be easily isolated.

Of particular interest in the above excerpt is the mention of the 'word gap'. This is defined in the document as the difference in vocabulary between children who enter school with an expected vocabulary for their age and those who enter school with a vocabulary lower than expected for their age. The 'word gap' refers to the concept that varying linguistic inputs available to children based on socioeconomic status may lead to meaningful differences in vocabulary size, which may impact academic outcomes (B. Hart & Risley, 1995; Hindman et al., 2016; Logan et al., 2019; Oxford University Press, 2018; Quigley, 2018). Put simply, poorer children may hear fewer words and develop smaller vocabularies. If these variables do affect the communicative experience of a child and are related to language outcomes, then "an integrated account of what in children's environments makes language acquisition possible while also making language outcomes variable" (Hoff, 2006, p. 58) is key in examining how human environments support language learning.

The source of the 'word gap' idea is a book entitled *Meaningful Differences in the Everyday Experience of Young American Children* (B. Hart & Risley, 1995). This book has had a remarkable effect across global educational policy and its ideologies have been imported into the United Kingdom (Cushing, 2023) despite criticisms (Avineri et al., 2015; Baugh, 2017; Burnett et al., 2020; Cushing, 2023; Dudley-Marling & Lucas, 2009; P. Nation, 1995) of the methodologies and the deficit

thinking at the source of the 'word gap'. Section 2.1 of the literature review will interrogate the original Hart and Risley study and the research that followed it to examine whether there is any evidence of this gap in the UK.

I will now briefly explain the structure and background of this thesis. The original intention of this project was to examine lexical development in children over time using two methods: tracking vocabulary size and growth by testing schoolchildren in person at regular intervals, and examining children's writing using the Oxford Children's Corpus. It was assumed that schools would remain open for the entirety of the project's length and that travel to Oxford would be feasible to enable collaboration on examining their corpus. However, six months after I commenced work and two after I had collected my first round of vocabulary data in March 2020, COVID-19 pandemic lockdown restrictions were put into place. While continuing the vocabulary testing was not logistically possible, the school who had offered to be a collaborative partner on this project was still keen to assist in other ways. The situation was challenging, but presented a unique opportunity to examine lexical development in a new context. The corpus was still examined, but to supplement the lost vocabulary work, several teachers agreed to participate in semi-structured interviews, providing a rich and interesting insight into how the pandemic had affected their teaching. This thesis presents the results of this adapted programme of research.

Following this introduction is the literature review. We begin by considering the word gap and its path to England's national curriculum, examining the emergence of this linguistic ideology in language policy in the UK (Chapter 2.1). We then examine prior attempts to measure vocabulary development in children, focusing on (a) vocabulary testing (Chapter 2.2) and (b) qualitative corpus linguistic approaches to evaluating children's writing (Chapter 2.4). Chapter 2.3 presents some background information on written language development in children.

Three experimental chapters follow the literature review. Vocabulary size testing is one method to estimate how many words children know, but the vocabulary sizes of school-age children using such tests are underreported. One reason for this is that few vocabulary tests exist that can give an accurate vocabulary size for L1 English children, especially for those who are pre-literate. These tests, their limitations, and what the literature says about vocabulary sizes of children will be discussed in Section 2.2 and Chapter 3. Chapter 3 is an examination of vocabulary size in children. It includes the data that were collected before the onset of the pandemic in March 2020, and discusses the merits and drawbacks of vocabulary size testing in L1 English children. Chapter 4 presents data collected from interviews with schoolteachers in lieu of the continuation of the vocabulary testing. Teachers were interviewed about the impact of the pandemic on their teaching and on their students' learning, with a focus on what they felt were the driving factors behind their pupils' vocabulary development during the pandemic. The impact of the pandemic on education is as yet unknown, but this thesis contributes to this research. Chapter 5 presents an analysis of the Oxford Children's Corpus, a collection of children's writing in one genre (short story writing).

In Chapter 6, these three disparate pieces of work are brought together to demonstrate the complexities involved in future research on vocabulary development in children in the UK, contextualised in the light of the COVID-19 pandemic. The key results are discussed and the idea of the word gap is re-examined in light of these results. Then, I discuss the idea threaded throughout all the chapters of this thesis—language policy—and reflect on how linguistic ideologies are made practice.

This thesis is influenced by critical discursive approaches to language policy. These approaches promote gathering and triangulating many data sources to show how language policy does not just happen in statutory documents, but is reproduced "across multiple layers of language policy interpretation and appropriation" (Johnson, 2012, p. 119). There are three levels to analysing

language policy: 1) the micro-level (e.g., teachers); 2) the meso-level (e.g., school management); and 3) the macro-level (e.g., government-produced curricula) (Cushing, 2020). These levels interact with each other to design, implement and enact policy in the UK. It is through these "overt and covert mechanisms" that those in authority (for the most part) control language spaces, as explained by Elana Shohamy in *Language Policy: Hidden agendas and new approaches*:

> the study of LP [language policy] should not be limited to formal, declared and official policies but rather to the study of the powerful mechanisms that are used in most societies nowadays to create and perpetuate "de facto" language policies and practices. The mechanisms discussed in this book include language education policies, language tests and language in the public space; it is through these policy devices that "real" policies are created. (Shohamy, 2005, p. xvi)

This thesis reflects these mechanisms: (1) Chapter 3 dissects a language test, considering their use and value in measuring vocabulary size; (2) Chapter 4 looks first-hand at how learning changed during the pandemic and how language education policies had to change, and considers the role of teachers and parents in children's vocabulary learning; (3) Chapter 5 considers a corpus of writing produced by children across the UK written for a public short-story writing competition, one way in which ideas about children's language are brought into the public space.

The title of this thesis poses a question: is there evidence of a vocabulary gap in the United Kingdom? The answer to this is not a straightforward yes or no. The longer answer is that the empirical studies conducted in this thesis did not allow any conclusions on this topic, and by the end of the thesis we will see how the originally designed studies never would have been able to answer this question, because researching the word gap is far more complicated than it might seem at first glance. Furthermore, prior research (to be introduced and discussed in Section 2.1) calls into question the very idea of the gap, highlighting that the word gap is complex and misunderstood. So,

while originally this research project was designed to provide evidence for the word gap, it did not yield these intended insights.

Although this project does not (and would not have been able to) show evidence of a gap in the UK, it regardless makes useful contribution to research and practice. First, this work presents several methodologies which can be used in vocabulary research, including vocabulary testing, corpus linguistics, and qualitative interviews. I highlight how important it is to explicitly define and operationalise concepts in linguistics research prior to commencing research. Second, this thesis presents literature showing how the word gap is a linguistic ideology that has been perpetuated throughout the United Kingdom and elsewhere despite limited, often conflicting evidence. I shed light on how complex the topic of the word gap is and how methodologically difficult it is to measure. It is hoped that this work will encourage a more nuanced view of the word gap, especially as concepts like this have far-reaching consequences for vocabulary teaching and learning.

2. Literature Review

Language is one of the most complex tools that humans possess. However, despite its complexity, language is acquired easily and naturally by typically developing children, even though acquiring a second language in adulthood presents difficulty for many. In this thesis, the focus is vocabulary development in L1 English-speaking children. Researching this is complex. Children learn words at different rates due to both individual and environmental differences, as well as due to stages in their language development. They acquire new words through parental and caregiver input, interaction with siblings and peers, reading, watching television, and school.

Answering the question of how children learn words involves the consideration of several complex interrelated issues, with the first a consideration of child development and language acquisition in general, as opposed to word learning in particular. Language is only a small part of the overall cognitive domain that changes over one's lifespan. Development is a multidirectional, multidimensional phenomenon that is influenced by both genetics (nature) and environment (nurture) (Paris et al., 2019). All children go through clear linguistic developmental stages independent of the language they are learning, and children also pick up grammatical features of language without explicit teaching. This happens in all typically developing children despite differences in inputs varying in quantity and type. Although children develop at different rates, there are some commonalities: for example, comprehension always precedes production, then they proceed in parallel, with production always trying to keep up with comprehension (Steinberg & Sciarini, 2006). The stages of written language development will be examined in Section 2.3.

Child language researchers have spent decades studying these *similarities* in the stages of language development, and language *variation* only became a key area of research in the 1980s, as researchers began to study larger samples of children and children from diverse language communities and varied cultural and socioeconomic groups (Goldfield et al., 2024). Research on

vocabulary acquisition has been complicated (Gleason, 2005) by methodological differences in measuring both comprehension and production. Methods to measure vocabulary size in children have ranged from diary studies, different instructional tasks, and picture-based tests (McDaniel et al., 1998). Section 2.2 will explore further specific methodological problems in measuring vocabulary in its discussion of vocabulary size tests. Another way to measure vocabulary development is analysing the words children use when they write. Quantitative corpus analysis approaches are particularly useful here. One systematic literature review identified 104 studies (published between 1945–2015) on language development in the *writing* of children in Anglophone countries, finding 66 studies on syntax, 45 on vocabulary, 19 on cohesion, and none on formulaic language (Durrant et al., 2021). This thesis will examine quantitative approaches to analysing children's written language in Section 2.4, which will present some of the literature on quantitative approaches to analysing children's writing, and Section 5, which presents a lexical investigation of the Oxford Children's Corpus, a large corpus of written fiction writing from children across the UK.

For now, we turn our attention back to the 'word gap' mentioned in the introduction. This word gap is a separate, but related, idea to the 'achievement gap' prevalent in education reform research and practice. The achievement gap is understood as the difference in national achievement tests between various student demographic groups, with minoritised or racialised groups performing worse (S. Anderson et al., 2007; Cabral-Gouveia et al., 2023; Hanushek et al., 2022). The achievement gap is often referenced in an American context as it is the basis for the controversial No Child Left Behind Act, which sought to narrow the achievement gap in the United States by creating common expectations for all (i.e., standards-based education reform). This act received wide bipartisan criticism in the US due to its penalisation of schools which showed no improvement in standardised test scores (C. Turner, 2015). This thesis will not consider the achievement gap as a whole, and instead will focus purely on the word gap.

What is the word gap? How do we measure it? The answers to these questions are complicated. We will start with the source of the idea of the word gap—the difference in vocabulary knowledge between advantaged and disadvantaged populations—and then go on to examine how the word gap discourse has been perpetuated in language policies across the globe. Examining language policy involves a consideration of how groups and individuals attempt to control and manipulate language in order to promote political, social, economic, and personal ideologies (Shohamy, 2005). It would be easy to think of language policy as something contained in policy documents created by ministries of education, but governmental bodies are only a small part of how language policy becomes language practices. Teachers and students are also key mechanisms by which ideologies are sustained and language spaces are controlled.

Before we begin, it is relevant to briefly summarise the state of English language policy in the United Kingdom. It is important to have some contextual understanding of how education is provided in the United Kingdom in order to understand how linguistic ideologies, especially deficit-based discourses, have been perpetuated throughout the preceding decades to emerge in their current form in current educational policy documents (all occurring at the macro-level of language policy).

Critical discursive approaches to language policy promote gathering and triangulating many data sources to show how language policy does not just happen in statutory documents, but is reproduced “across multiple layers of language policy interpretation and appropriation” (Johnson, 2012, p. 119). To reiterate from the introduction, these layers include 1) the micro-level (e.g., teachers); 2) the meso-level (e.g., school management); and 3) the macro-level (e.g., government-produced curricula) (Cushing, 2020). These levels interact with each other to design, implement and enact policy. Overt and covert mechanisms help those in authority control language spaces, and interact at all three levels (the micro, meso, and macro) to convert linguistic ideologies to language

practices. I will return to a discussion of these levels in Chapter 6 when I discuss the results of this thesis through a policy lens. For now, it is enough to remember that language policy happens in disparate points in space and time, reproduced not only when governments mandate curricula, but also when teachers idly correct their students in class, or when a headteacher signs off on posters which inadvertently police and regulate language use (Cushing, 2020; Johnson & Ricento, 2014; Shohamy, 2005).

Education in the United Kingdom is devolved, with the four nations (i.e., England, Wales, Scotland, and Northern Ireland) each responsible for providing their own national curricula. Throughout this thesis literature and data will be contextualised with reference to the national curriculum where relevant, and this refers only to the national curriculum for England. The nations will not get equal representation in this thesis. England will be used as the primary source and for examples throughout most of this thesis due to the higher amount of empirical research and publications from organisations such as Ofsted (the Office for Standards in Education, i.e., the school inspectorate in the United Kingdom) and the Department for Education (the ministerial department responsible for education in the United Kingdom).

In England, the majority of schools follow the national curriculum, which was introduced by the Department for Education in England in 1988 in order to standardise classroom teaching (Goodwyn, 2014). This is statutory (i.e., mandatory) guidance, and thus schools have a legal obligation to follow the curriculum. Wales, Northern Ireland, and Scotland have separate educational policies and systems under separate, devolved governments. In April 2021, the Welsh government passed the Curriculum and Assessment (Wales) Act (2021), establishing a new framework for the curriculum for pupils in Wales. This new Curriculum for Wales, which is being gradually phased in at the time of writing, represents a significant departure from the much more prescriptive guidance of the national curriculum in England. Control of curriculum and assessment

design and implementation is given back to schools, with head teachers responsible for designing and implementing the curriculum in their school. The new Welsh Curriculum promotes language and literacy as a tool that can help students become informed, global citizens, and that acknowledges that children progress at different rates (Welsh Government, 2020). Literacy is enshrined as a "mandatory cross-curricular skill" in the Literacy and Numeracy Framework. Multilingualism is emphasised alongside the importance of valuing all languages and cultures.

The English national curriculum was implemented at a time of educational reform, where education was moving towards a system of centralised control rather than the fairly democratic system which had developed throughout the twentieth century, where education was largely under local authority control. The curriculum has been criticised for curtailing teacher autonomy (U. Clark, 2001), for the increased prevalence of language policing in schools, especially of non-standardised L1 language (Cushing, 2020), and for its nationalistic ideologies (Yandell, 2017). In this thesis, we will also see how the word gap is linked to language policing and raciolinguistic bias.

2.1 The Word Gap: Ideology to Policy

Meaningful Differences in the Everyday Experience of Young American Children (B. Hart & Risley, 1995) presents research conducted by Betty Hart and Todd Risley (referred to in this thesis as H&R) in the 1980s and 1990s. Interested in examining the differences in the linguistic experiences between poor and middle-class children, H&R designed a study to capture the language that children heard at home. This was one of the first attempts to capture the naturalistic home language environments of children in the United States.

The study was labour-intensive: over a period of 2.5 years, the research team collected over 1,300 hours of audio which took an additional 4.5 years to transcribe, code, and analyse. The participants were 42 families in Kansas City. The cohort was stratified by income: 13 upper-

socioeconomic status (SES) families (1 Black, 12 White), 10 middle-SES families (3 Black, 7 White), 13 lower-SES families (7 Black, 6 White), and 6 families receiving welfare (governmental benefits) (all Black). Each family was visited once a month, where researchers recorded one hour of talk and interaction with each family. This one hour per month of data was extrapolated to generate daily count estimates.

A summary of some of H&R's conclusions are as follows:

1. Children from families on welfare heard about 616 words per hour, while those from upper-SES families heard 2153 words per hour;
2. This means that children of upper-middle-class families hear 30 million more words by the age of three than children of low socioeconomic status families;
3. By age three, children from upper-middle-class families had produced over 1,200 different words, whereas children from welfare-dependent families had only 600 different words in their productive vocabularies;
4. Parents talked to their children differently, and children from families on welfare were subject to more instances of negative reinforcement;
5. Children from the professional families heard six encouragements for every discouragement.

The main takeaway from Hart and Risley's work (that there may be diverging linguistic inputs available to children depending on class) inspired a raft of further research and had wide-reaching policy implications aimed at reducing this gap. The thinking went as so: Language acquisition is dependent on an interplay between genetic and biological factors and socio-cultural contexts and environmental factors. One confounding factor is socioeconomic status. Socioeconomic status, defined in this thesis as an individual's access to economic resources and social status, is often linked to language development (as well as to general development and health, among other things). Research on the impact of socioeconomic status on early language

development is complex due to methodological considerations, such as how socioeconomic status is operationalised (Goldfield et al., 2024). Some conclusions have found a link between socioeconomic status and the amount of lexical input (i.e., words) that children hear, and thus how many words they learn. Put simply, some argue that poorer children learn fewer words than richer children because they *hear* fewer words. The H&R publication is one source of support for this argument, and other supporting findings are summarised by Goldfield et al., p. (2024, p. 267) in the most recent edition of *The Development of Language*:

> First, children from low-SES households, on average, learn language more slowly than children from higher-SES households, meeting vocabulary and other milestones at a later age (e.g., Arriaga et al., 1998; Farkas & Beron, 2004; Fernald & Marchman, 2012). These differences emerge by 18 months and are reflected not just in vocabulary size but also in processing speed and efficiency (Fernald et al., 2013). Second, parents with more education and higher incomes, on average, talk more to their children, are more responsive to their children's language, and encourage children to talk more, than parents with low levels of education, with low incomes, and/or experiencing high levels of life stress (e.g., Hoff et al., 2002; Rowe, 2008). Third, differences in children's language development, in particular speed of vocabulary acquisition, are strongly related both to quantity (e.g., Hoff, 2003; Huttenlocher et al., 1991; Pan et al., 2005) and to quality of input (Rowe, 2012; Weizman & Snow, 2001). Quality can be measured along different dimensions: conversational responsiveness, lexical variety and syntactic complexity, and conceptual challenge (Rowe & Snow, 2019); all of these are likely to vary with parental education and income.

Hart and Risley's research drew attention to variation in the linguistic environments of different social classes and captured the minds of educators, researchers, and laypeople. The effect on policy in both the United States and the United Kingdom has been enormous. Given the

understanding that there is an association between English language skills and academic achievement, this variation was identified as a key area which could be targeted for improvement. Public health and educational interventions aimed at reducing the word gap across the globe followed (Greenwood et al., 2017). After all, this issue seemed actionable: encourage parents to speak more to their children, or to speak differently. In the US, federal investments included Head Start and Early Head Start, and local interventions included Providence Talks in Rhode Island, Reach Out and Read Out in Boston, and the Clinton Foundation's Too Small to Fail. In the UK, from 2017, government initiatives worth over £100 million have been funded with the aim to 'close the gap', including several schemes targeted at changing the speech of low-income and racially minoritised children by encouraging parents to talk more or differently with their children (Cushing, 2023). Some of these interventions and their outcomes will be further considered in Section 6.

Some have levelled critiques against the original study in regard to its methodological and theoretical approaches. The main theoretical criticism of the original study and the subsequent research supporting its claims is that this research, by its nature, promulgates a deficit language model, perpetuating discriminatory views against non-mainstream ways of speaking (Avineri et al., 2015; Baugh, 2017; Cushing, 2023; Dudley-Marling & Lucas, 2009; P. Nation, 1995; Sperry et al., 2019b; Yandell, 2017). It is argued that the sample of 6 welfare families, all Black, and 13 professional families, 12 of whom were White, "reinforce[s] harmful stereotypes that conflate poverty and race" (Dudley-Marling & Lucas, 2009, p. 364). To make this research evidence of a supposed 'word gap' between privileged and less privileged groups blames "families living in poverty for not living up to the standards of the middleclass rather than on the ideologies and practices of institutions responsible for the material conditions of families living in poverty" (Kuchirko, 2019, p. 543). Dudley-Marling and Lucas argue that "strong claims about language deficiencies in poor children and their families based on the Hart and Risley study are unwarranted" and that the

"uncritical acceptance" of their findings highlights a trend in some educational policymakers, practitioners, and researchers to pathologize the language and culture of poor students and their families (Dudley-Marling & Lucas, 2009, p. 362).

For example, one of the differences observed by Hart and Risley was that the poorer families (who were all Black) tended to use directives ("Pick up the toys") with their children rather than the indirect questions ("Why don't you pick up the toys for me?") used more commonly by the upper-SES families. Hart and Risley go on to assert that these direct requests are examples of a "prevailing negative attitude" (B. Hart & Risley, 1995, p. 177) that would negatively affect children's linguistic development, and which would be difficult to undo. As there is no evidence that the welfare parents and their children considered these interactions negative, it is arguable that Hart and Risley construed the interactions between the upper-SES families as higher quality because they reflected their own values (Dudley-Marling & Lucas, 2009).

It is unclear whether the families engaged in comparable activities while they were being observed. Parents were asked to choose when they would be observed, with many choosing mealtimes. It is possible that upper-SES, professional parents chose to be observed during activities containing more language opportunities than poor families (Dudley-Marling & Lucas, 2009). Hart and Risley also gave no consideration that their observers audiotaping the one hour of verbal interactions between parents and children may have affected the language that the parents were using with their children. Research shows that observers can have a significant effect on child–caregiver interactions (Zegiob et al., 1975) and the methods used by Hart and Risley do not reflect current approaches in ethnographic research which places importance on local ecologies, cultural meaning systems, and embracing normative language practices (Sperry et al., 2019b). Gilkerson et al. (2017) carried out a near-replication study utilising discreet digital recording devices with 329

families. These authors found that the "gap" between high-income and low-income children was around 4 million words by the age of four, not the 30 million put forward by Hart and Risley.

In their attempt to replicate Hart and Risley's findings, Sperry et al. (2019b) highlight that dyadic interactions, such as the interaction between a primary caregiver and a child, are socioculturally defined, and do not exist in many cultures. They emphasise that

> the psycholinguistic tradition has produced strong evidence that young children can learn vocabulary from directed speech *and* from overheard speech. […] speech directed to children predicts later vocabulary growth and language outcomes in school and suggests that different features of directed speech may have different effects as children get older. (Sperry et al., 2019b, p. 4)

When Sperry et al. (2019b) widened the definition of speech in their near-replication of the H&R study to number of mean words said by *all* caregivers to a child, or to the amount of *all* ambient speech within a child's hearing, it was notable that the word gap between different social classes disappeared. In fact, in their study, poor and working-class communities showed an *advantage* in number of words heard. They argue that the approach taken by Hart and Risley ignored several important sources of the linguistic environment available to children, including speech directed at children from adults other than the primary caregiver as well as other speech that occurred around the children: H&R "defined and operationalised the children's vocabulary environment as speech directed to the child by the primary caregiver, excluding speech by other family members to the child as well as speech that the child overheard" (Sperry et al., 2019b, p. 2).

This focus on child-directed speech is not unusual. There is a great body of research showing that child-directed speech is a source of linguistic input that can boost lexical and syntactic development (Fernald et al., 2013; Hurtado et al., 2008; Huttenlocher et al., 2007). However, other research suggests that young children in societies where they are seldom spoken to (i.e., receive less

child-directed speech) nonetheless attain the same linguistic milestones and do not suffer any apparent delays in their linguistic development (Casillas et al., 2021; Hellwig, 2022). Brown and Gaskins (2014) argue that the variation in a child's speech environment leads to different strategies for attending to, and learning from, speech. This idea supports evidence that shows that different cultural practices yield different developmental strategies across many domains (Goodnow et al., 1995; Miller et al., 2012).

Sperry et al. go on to explain how H&R discouraged adult–adult conversation and eliminated adult talk that was not directed to the child from their analysis. By doing this, Hart and Risley did not capture features of family life and their associated types of talk, such as multiparty and bystander talk (Sperry et al., 2019b). The authors conclude that their findings do not support Hart and Risley's claims and clarify that the individual variation within and between families and communities in particular social classes may be so great that it obscures variation across classes. H&R did not consider that verbal experience is not uniform in any social class and that there is much more to the verbal environment than child-directed speech from a primary caregiver.

In their critique of Sperry et al.'s (2019) paper, the authors agree that they "do not support the deficit perspective that all low-income parents do not talk enough with their children" (Golinkoff et al., 2019, p. 7), but argue that denying the existence of this gap has serious consequences. This is because research into the word gap has led to several interventions aimed at building language and improving outcomes in school, and that language is key to children's school achievement. In a follow-up response, Sperry, Sperry and Miller defend their position that the influential claim made by H&R deserves more scrutiny, and that research is moving toward re-thinking our understanding of the nature of young children's language environments, which should not be defined only in terms of speech directed to the child by the primary caregiver (Sperry et al., 2019a). I will not delve any further into this specific debate, but will affirm that I agree that the idea of the gap warrants more

scrutiny. This is especially pertinent in the United Kingdom, where there has been no comparable research on vocabulary differences between different social classes, and yet the vocabulary gap is mentioned in research commissioned by the government on the English national curriculum, suggesting that this ideology is commonplace in the Department for Education.

As we consider if there is evidence for a word gap in the United Kingdom, it is worth bearing in mind best practices related to linguistic research in this area. The British Association for Applied Linguistics, in the most recent edition of *Recommendations on Good Practice in Applied Linguistics,* state that when conducting research "it is necessary to consider the effects of research on all groups, including those that are not directly involved at the time; for example, the potential for research approaches to reinforce assumptions about minoritised language[1] communities" (BAAL, 2021, p. 37).

This directly relates to criticisms that the word gap only pathologizes the language of socioeconomically disadvantaged groups by making the vocabulary of the privileged the yard stick by which everyone else is measured. It is important that linguistic research is informed by sociocultural theories of language which consider the strengths and diversities of non-dominant communities, taking the language practices of different groups on their own terms rather than only being seen in reference to the language of dominant groups. Cushing (2023, p. 306) gives us a succinct definition of the raciolinguistic ideologies which underpin the word gap:

---

[1] A minoritised language is a language which is marginalised, suppressed, or banned. Using *minoritised* rather than minority highlights the processes which value or devalue a language, recognising that many so-called "minority" practices are normative in the groups which practice them (Flores & Rosa, 2015).

> Raciolinguistic ideologies represent beliefs about language tethered to European colonialism and its ongoing legacies, within which low-income and racialized speakers' language practices are perceived as deficient, incomplete, and indeed, full of gaps when compared against the language practices of the idealized white middle-classes.

The deficit discourse has had a recent resurgence in England's education policy (Flores & Rosa, 2015), which is "evident in discourses, funding and classroom materials related to the so-called 'word gap' or 'vocabulary gap'" (Cushing, 2023, p. 306). Cushing (2023) describes how politicians have used the word gap to place the blame for attainment differences between children from different socioeconomic differences at the feet of their parents. This reflects similar attitudes from the 1950s, where increasing diversity in the classroom led to a prevailing attitude that it was up to teachers to fix working-class educational underachievement by giving these pupils the language they 'needed' to succeed in education (i.e., standard English) (Hardcastle & Yandell, 2018). Concerns over the linguistic deficiencies in the homes of underperforming children in British schools were again revived in the 1980s, but this time aimed at Black students, primarily from African-Caribbean backgrounds (Hardcastle & Yandell, 2018). It seems that now the deficit discourse has manifested once again under the guise of this supposed 'word gap'.

So, how was the word gap ideology imported into the UK? At the time of the original H&R study, the 'word gap' or 'vocabulary gap' concept was readily accepted by researchers, teachers, journalists, parents, and politicians. From there, the word gap discourse affected educational policies across the globe; the gap has even been framed as a public health crisis (Cushing, 2023). In the United States, the original Hart and Risley study was so influential that it has been cited in Congressional hearings as reason for the need for evidence-based early childhood programmes to reduce the gap (Dudley-Marling & Lucas, 2009).

After its emergence in the United States, other countries began to take note. Cushing (2023) provides a detailed timeline of how and when word gap ideologies were imported into England, showing how the deficit discourse has been weaponised by politicians and policymakers, most of whom are white men (e.g., Michael Gove, Nick Clegg, Nick Gibb), to advance their political agendas. For example, in 2015, Nick Gibb (a conservative MP and an education minister for England at the time) contributed to a set of essays for Policy Exchange, a UK-based think-tank which, despite claiming bipartisanism, is heavily linked to the UK right (A. White, 2013). In this essay, entitled *How E.D. Hirsch Came to Shape UK Government Policy*, Gibb discusses the influence of American outcomes-based education theories on his own approach to education reform (Gibb, 2015). In this essay, he cites the Hart and Risley study as proof of an "incontrovertible" divide between children who arrive at school with more language compared to those who do not, linking this to worse outcomes. He suggests that the knowledge-based curriculum put forward by the Conservatives, and which has been in place in England since 2013, is the key to fixing this divide.

Leaving aside criticisms of this approach to educational reform for the moment, it is notable that nowhere does Gibb engage with or reference any of the criticisms of the word gap. He presents this single study as proof of "inequality in terms of mental architecture" which "provides a clear case for a knowledge based curriculum at an early stage" (Gibb, 2015, p. 14). Gibb goes on to highlight that the language gap ideology is "an argument which falls upon receptive ears across the political spectrum" (p. 14). This is backed up by reference to Hirsch's idea of intellectual capital and the idea that if children arrive to school with more language, they are inevitably bound to do better. This idea of 'the rich get richer' is referred to as the Matthew effect in language acquisition, after Matthew 25:29: "For whoever has will be given more, and they will have an abundance. Whoever does not have, even what they have will be taken from them." (*New International Version*, 2011). This essay as well as the explicit reference to the word gap in England's national curriculum are examples of the

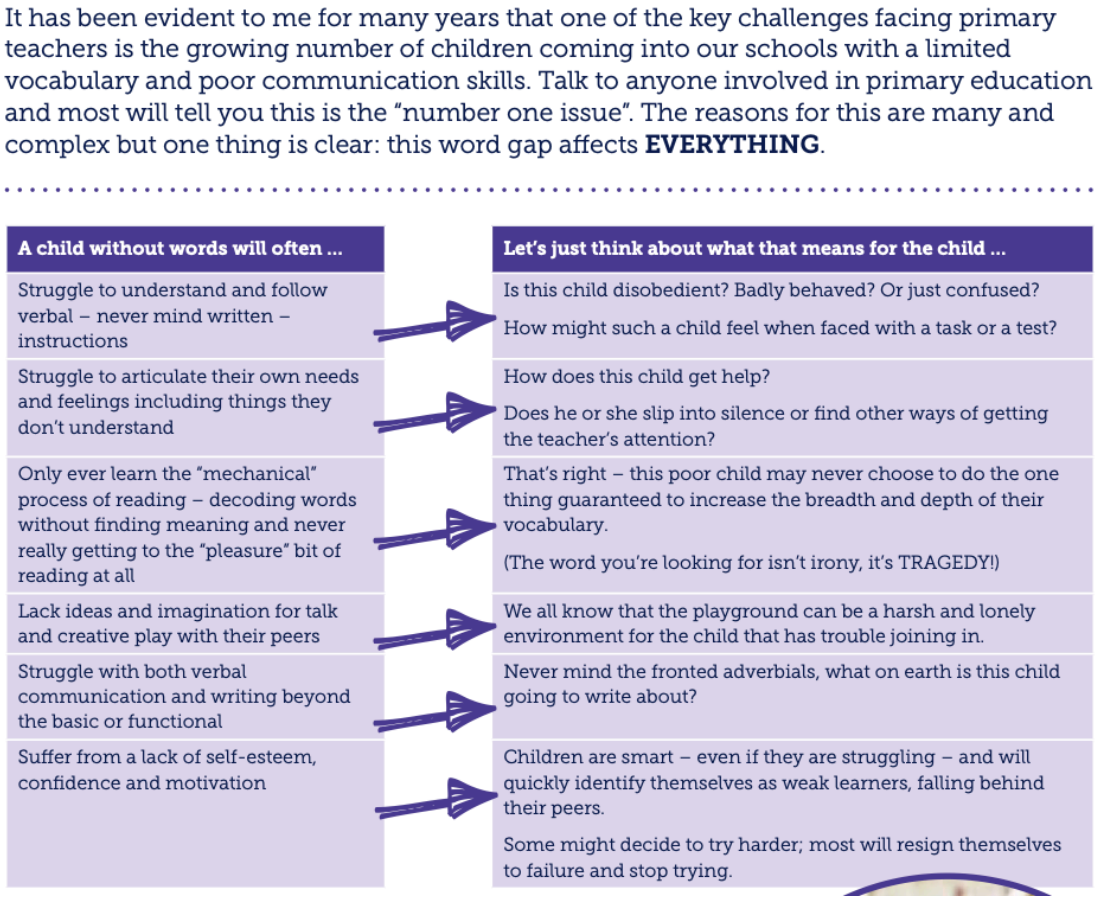
'layers' of language policy and show how ideologies can become fossilised into state-controlled educational policies (Cushing, 2020; Shohamy, 2005). Uncovering these hidden ideologies is key to improving language education.

Non-governmental institutions can also influence language ideologies and policies (the meso-level). In the United Kingdom, Oxford University Press have been a large driver of the word gap discourse, releasing regular publications on the scope of the issue in Britain. We will now briefly examine several of these reports.

In their report *Why Closing the Word Gap Matters* (Oxford University Press, 2018), the Children's Language team of Oxford University Press (OUP) present 'market research' on the word gap in schools. The report opens with the claim that "without enough language–a word gap–a child is seriously limited in their enjoyment of school and success beyond" (Oxford University Press, 2018, p. 2). Their evidence for this claim is a survey distributed to 1,313 teachers (840 secondary school teachers and 473 primary school teachers) between December 2017 and January 2018. The statistics presented are stark: 40% of pupils lack the vocabulary needed to access their learning; over 60% of teachers think the gap is increasing; 49% of Year 1 pupils have a limited vocabulary to the extent it affects their learning; over 90% of teachers believe that lack of time spent reading for pleasure is a root cause of the word gap; and so on. These facts are presented plainly, sensationalised into the "number one issue", and linked to outcomes such as difficulty following what is going on in class, slower progress in subjects other than English, and even worse mental health and life chances. Figure 2.1 presents an example of the ideas put forward by this report.

Figure 2.1

Excerpt From 'Why Closing the Word Gap Matters' (Oxford University Press, 2018, p. 11)



It has been evident to me for many years that one of the key challenges facing primary teachers is the growing number of children coming into our schools with a limited vocabulary and poor communication skills. Talk to anyone involved in primary education and most will tell you this is the "number one issue". The reasons for this are many and complex but one thing is clear: this word gap affects **EVERYTHING**.

A child without words will often ...	Let's just think about what that means for the child ...
Struggle to understand and follow verbal – never mind written – instructions	Is this child disobedient? Badly behaved? Or just confused? How might such a child feel when faced with a task or a test?
Struggle to articulate their own needs and feelings including things they don't understand	How does this child get help? Does he or she slip into silence or find other ways of getting the teacher's attention?
Only ever learn the "mechanical" process of reading – decoding words without finding meaning and never really getting to the "pleasure" bit of reading at all	That's right – this poor child may never choose to do the one thing guaranteed to increase the breadth and depth of their vocabulary. (The word you're looking for isn't irony, it's TRAGEDY!)
Lack ideas and imagination for talk and creative play with their peers	We all know that the playground can be a harsh and lonely environment for the child that has trouble joining in.
Struggle with both verbal communication and writing beyond the basic or functional	Never mind the fronted adverbials, what on earth is this child going to write about?
Suffer from a lack of self-esteem, confidence and motivation	Children are smart – even if they are struggling – and will quickly identify themselves as weak learners, falling behind their peers. Some might decide to try harder; most will resign themselves to failure and stop trying.

Here, the word gap is linked to "everything", including struggling to understand instructions, a lack of creativity, and low self-esteem. A description of methods is not included in this report (its audience is educationalists and policymakers, not academics) but OUP include examples of the sorts of things they found out in their survey, such as (Oxford University Press, 2018, p. 8):

- 60% of secondary teachers surveyed (of 840 secondary teachers who took part) reported that the number of students who have a limited vocabulary is either increasing or significantly increasing.

- 80% of secondary teachers surveyed reported that students who have a limited vocabulary find reading national test papers very or extremely challenging and 79% felt that they achieved worse results in national tests.
- 80% of secondary teachers surveyed reported that the word gap leads to lower self-esteem, 65% report that this has a negative impact on behaviour, and 82% report that students are less likely to stay on in education.

What constitutes a 'limited vocabulary' is not defined, and there is little to no acknowledgment of the limitations of this research, such as that this is all based on teachers' perceptions, not actual evidence of results. Their results are "corroborated" by "numerous studies" and "abundant evidence" (Oxford University Press, 2018, p. 5) showing that the input that children receive is the key determinant of how they develop language. However, the only study referenced is the original Hart and Risley study, critiqued above. These allusions are "used to craft a narrative of scientific robustness and academic legitimacy" (Cushing, 2023, p. 319).

In 2020, a follow-up report built on this research (Oxford University Press, 2020), focusing on the impact of the COVID-19 pandemic. This report exemplifies the way in which the gap discourse shifted during the pandemic, presenting the 'COVID language gap', the idea that children missed out on key input from stay-at-home closures, and this exacerbated the already present (and growing) vocabulary gap in UK schoolchildren. In this report, OUP link the word gap to the failure of parents to talk to their children during the pandemic. Cushing (2023) highlights the lack of engagement with socioeconomic structures and inequalities that were exacerbated during the pandemic and also criticises this report for its lack of adequate references to academic literature, lack of attention paid to the history of the word gap, and lack of a critical analysis of the problems at the heart of deficit-based discourse. It is clear to see why these critiques have been levelled against the report. The documents produced by Oxford University Press are underpinned by deficit-based ideologies that

link poor vocabulary with lowered chances of success in school and, subsequently, in life. These are annually perpetuated, with other reports claiming that the pandemic has exacerbated the gap, and that teachers tend to agree that children with more support for vocabulary development at home perform better in school (Oxford University Press, 2023). These claims are presented without interrogation of what this really means or examination of whether this is correlation or causation.

We will return to examine how the word gap ideology has been reproduced by Oxford University Press in the United Kingdom in Section 6. For now, we consider the following question: How would we measure the words children in the UK know and thus determine which students are "word-rich" compared to those who are "word-poor", as the Department for Education highlights as a key issue in education in the UK? Some methods in applied linguistics which can be used to examine this include direct vocabulary size measurement and quantitative corpus linguistic analysis of children's writing. The remainder of this literature review will examine these methods.

2.2 Vocabulary Testing

Being able to 'measure the gap' involves some measure of being able to quantify it. One way to quantify this knowledge is through calculating a child's vocabulary size. However, this is a complex task. Measuring vocabulary size is tricky, involving detailed considerations of issues such as what is a word, the best method for measuring vocabulary size, choosing the words that should be tested, etc. Researchers have differed in the way they choose representative samples of words and the way in which they measure knowledge of those words (Brysbaert et al., 2016; P. Nation, 1993; P. Nation & Anthony, 2016). There have been several attempts at calculating the vocabulary sizes of English-speaking adults, but results vary greatly, with estimates of 10,000 to over 200,000 words (see Brysbaert et al., 2016; further discussion of the vocabulary sizes in Section 3). This variation is due to methodologies employed in the development of vocabulary tests, not lack of interest or research (P.

Nation & Anthony, 2016, p. 365). This section will explore some of the decisions that have to be made when developing and administering a vocabulary size test. Specifically, we will consider these issues in the context of designing receptive vocabulary tests for children. Specific studies on vocabulary sizes in L1 English in adults and children will be presented in the literature review of Section 3.

Vocabulary size testing at all stages of human development is of interest to researchers to model the development of the human lexicon. Psychological testing often includes a verbal element: "verbal ability is probably the best single indicator of crystallised intelligence almost certainly because, in Western culture, education is essentially verbal" (Kline, 2000, p. 218). This is because verbal ability and word fluency have been linked educational and occupational success (R. C. Anderson & Freebody, 1981; Biemiller, 2011; F. J. Duff et al., 2015; Lee, 2011). In academia, vocabulary tests have been used for decades to calculate vocabulary sizes, but validation research still lags behind our understanding of various technical aspects of test design. There has recently been a call for greater rigour in test design and validation in the field of vocabulary testing (Schmitt et al., 2020). For the future of language testing research, it is important to provide thorough accounts of development procedures and explanations of how tests are developed and how constructs are operationalised, tested, and revised (O'Sullivan & Weir, 2011).

Bachman and Palmer, two pioneering American scholars in the field of language test design, state that "there is no such thing as a 'good' or 'bad' test in the abstract" and that "there is no such thing as the one 'best' test, even for a specific situation" (Bachman & Palmer, 1996, p. 6). They also highlight several misconceptions that exist in the field of language testing (and that still pervade the field to this day), such as believing that a perfect test exists, or having unreasonable expectations for a test. The purpose of a test is the most important thing to consider in the development process.

Vocabulary testing is carried out to achieve different aims and goals; the reason for vocabulary testing influences its design and implementation. Testing may be carried out frequently, such as to assess progress through a language, or it may be carried out periodically or even just once, such as for a university entrance exam or for vocabulary size measurement. Furthermore, tests may attempt to measure several aspects of language or just one underlying construct. All of these decisions will affect test design and implementation, and thus affect our interpretation of test scores. There is no perfect test, as a test that is suitable for one situation or group of learners will be unsuitable for others (Bachman & Palmer, 1996).

Tests can be viewed and used as tools of social power (Fulcher, 2009; Shohamy, 2001). The outcomes of tests often control decision-making for the allocation of potentially scarce resources, such as jobs or university places, and thus testing reveals social and political values. Tests are also symbols of success and achievement, evidenced by the very existence of a 'good' or 'bad' test score (Fulcher, 2013). To summarise:

> In most societies tests have been constructed as symbols of success, achievement and mobility, and reinforced by dominant social and educational institutions as major criteria of worth, quality and value. (T. McNamara & Shohamy, 2008, p. 89)

Researchers must follow several steps when designing a vocabulary test and must be clear about the purpose of a test and provide explanations and rationales for every decision made during the test development process (Schmitt et al., 2020). In their book *Language Testing in Practice: Designing and Developing Useful Language Tests* (1996), Bachman and Palmer put forward a framework for language test development and others have built on this to offer new frameworks for vocabulary test design and development (Chapelle, 2012; Weir, 2005). P. Nation & Coxhead (2014) summarise the steps that a vocabulary researcher can follow as such:

(a) Develop a standard and reliable unit of word counting that suits the goals of the test.

(b) Develop substantial word family lists that include virtually all the vocabulary to be tested.

(c) Use corpora to arrange the word families in the lists into frequency levels.

(d) Draw representative samples of word families from the lists so that no word frequency level is overrepresented or underrepresented in the samples.

(e) Decide on the test format and develop items for the tests.

This section will briefly address each of these steps in turn, in order to describe the process of creating research-based, methodologically sound vocabulary size tests that can be used with some confidence to gather data for research or practical purposes.

Finally, it is important to highlight that this section is only addressing receptive vocabulary, i.e., the words that people hear or read. Receptive vocabulary is only one dimension of vocabulary knowledge, that of breadth (Milton & Fitzpatrick, 2013; Read, 2000). Breadth relates to the number of words we know, while depth relates to how well we know them. Productive vocabulary items are words that can be actively produced when speaking or writing. Breadth and depth are best perceived as two axes of measurement which always work together (Durrant et al., 2021). Generally, it is assumed that receptive vocabulary knowledge is roughly twice that of productive vocabulary knowledge.

2.2.1. Definitions

To calculate the number of words in a document seems like a straightforward task: count the number of words. In modern text processing, English words are produced as strings of letters with a space on either side. But is that correct? How do we deal with inflected forms of words, i.e., do we count *swim*, *swam* and *swimming* separately? What about *swimmer*? Should numbers be considered words? Should *dog* and *dogs* be counted separately?

The variation in vocabulary size reports for L1 English speakers is due to how researchers have quantified the number of words in a language. This quantification depends on how one defines a word. This section will introduce several key definitions which will be used throughout this thesis. The explicit definition of these terms has often been the root source of many methodological problems in applied linguistics research. Common terms which must be clarified when undertaking vocabulary research include:

1. *Tokens* are the number of individual running words in a text, i.e., a string of letters with spaces either side, which may be repeated and counted more than once;
2. *Word types* are distinct tokens, counted only once;
3. *Alphabetical word types* are word types that consist only of letters, eliminating words that contain hyphens or other punctuation marks, not distinguishing between uppercase and lowercase letters;
4. *Lemmas* are the uninflected forms of a word (often called a *root*, *base*, or *stem*), including all the related word forms within one word class, e.g., *walk*, *walks* and *walking*, which are all forms of one verb;
5. *Word families* include all the related words that are not necessarily in the same word class, thus capturing all the word forms related to one concept e.g., *walk*, *walking* as well as *walker* and *walkable* (including verbs as well as a noun and an adjective).

The choice of whether to define words at the lemma or word family level is a key methodological consideration in lexical analyses and there is ongoing discussion on which lexical unit is most appropriate for research and pedagogy (D. Brown et al., 2021; Webb, 2021). Theoretically, the question is whether having knowledge of one item can predict knowledge of the other items in the group, i.e., if someone knows *walk*, can we assume they know *walkable*? Practically, grouping more tokens together will affect calculations, i.e., lower scores in measures of lexical diversity. The

main point when deciding on the unit of count is that it should be chosen carefully, based on the reason for the test being carried out and the participants, and it should be clearly articulated which unit has been chosen and why. If these issues are addressed, research will be both methodologically sound and comparable. Moving forward in vocabulary research, a more principled and systematic manner of defining a word is vital.

The use of different definitions will result in different measures of the overall size of a text and will affect the outcome of a vocabulary size test. This means that one of the first things to address when designing a vocabulary test is to choose a principled unit of count and explain why that has been chosen. The choice should be based on the research question and the group being researched (Schmitt, 2010; Schmitt et al., 2020). For example, when measuring first language acquisition in the early years, counting different word types may be the best count because the acquisition of these new forms in the early years is important. For adults, lemmas are appropriate, as we can assume they have mastered morphology and can use grammatical affixes with some confidence. Lemmas are a popular unit of count because they have a transparent definition and it is easy to see what is being measured (P. Nation & Anthony, 2016). In fact, Schmitt (2010) argues that lemmas may be the best unit of count overall as “they are unambiguous to interpret” (p. 193). Furthermore, using different units of count for receptive versus productive knowledge means that tests of vocabulary size may not be comparable. For this reason, Schmitt (2010) suggests that the lemma may be the best unit of count overall so that we can make and deliver directly comparable vocabulary size measures for both receptive and productive vocabulary. P. Nation (1993) argued that lemmas may be “too restrictive” (p. 26) for measuring receptive vocabularies, except for low-proficiency learners, but that lemmas may be a sensible unit of count for measuring productive vocabulary. Despite this, the lemma level seems to be the most appropriate as there is evidence that learners who can recognise one member of a lemma are usually able to recognise other members,

but that this is not the same for word families (D. Brown et al., 2020); using lemmas may also allow for greater precision and fewer assumptions about knowledge, especially in corpus linguistics where the tools used can easily find and count lemmas (Gablasova & Brezina, 2021). Webb (2021) argues that there is no 'lemma dilemma' due to the fact that the majority of software tools are lemmatised and using lemmas is more precise and requires fewer assumptions about the morphological and semantic knowledge of learners. This is the position that this thesis takes.

2.2.2. Word Lists, Corpora, Frequency and Sampling

The next few sections are related to the selection of target words in vocabulary tests. The selection of words must be explicitly outlined, with justifications provided, so that the procedures within could be replicated if necessary (Schmitt et al., 2020). Early vocabulary size measures used dictionary-based methodologies, but these methods have several issues, such as the difficulty of taking an even and representative sample (P. Nation, 1993; P. Nation & Anthony, 2016).

Frequency, i.e., how common a word is (*dog* is more frequent in the English language than *cochlear*), is one of the most important aspects of lexical knowledge as it affects acquisition, processing, and use of vocabulary (Schmitt, 2010, p. 63). Frequency is associated with writing and speaking proficiency (Kyle & Crossley, 2015; Laufer & Nation, 1995), text complexity (P. Nation, 2006), word processing time (see Brysbaert et al. (2018) for a recent review on the frequency effect on word processing times), and word recognition tasks (Brysbaert et al., 2016). It is generally assumable and has been proven that we learn more frequent words before less frequent ones, in both L1 and L2, with frequency affecting all aspects of language processing (Ellis, 2002; Leech & Rayson, 2014).

Frequency analysis consists of counting all the words in a text to work out how often they occur; frequency figures can then be generated and words can be grouped into frequency 'categories'. These 'categories' are usually referred to as bands, and split per one thousand words.

For example, frequency bands could be the 1,000 most common words in English (1k band), followed by the next most common 1,000 words in English (the 2k band), and so on. It is common now to design vocabulary tests using frequency-based lists.

Frequency-based approaches have several limitations. The words in the test must be relevant for the people being tested, i.e., using low-frequency, difficult words for young children will result in low scores on a test, and using high-frequency vocabulary only in a test for adults will result in artificially high scores as they reach a ceiling. For example, using word frequencies from a corpus of English spoken/written by adults, e.g., the BNC (British National Corpus), will result in a test that is not suitable for young children, as some words being tested may not be known to them and the child would be evaluated in comparison to adult language. Durrant & Brenchley (2019b) highlight that this can be a deliberate approach if the vocabulary being measured is being viewed in terms of educational aims, i.e., that children are aiming towards adult language. The alternative would be to test children based on wordlists that contain vocabulary children are likely to know, especially as children may learn particular vocabulary items from semantic categories that are classed as infrequent in everyday, adult English, such as wild animals (e.g., *zebra, lion, giraffe*). A vocabulary test could thus be designed using a corpus of children's literature, school textbooks, or a spoken corpus of children's own language, such as the Oxford Children's Corpus of Reading (Banerji et al., 2013; Wild et al., 2013). This would evaluate children's language in comparison to children, which Durrant and Brenchley refer to as a "backward-facing reference point" (2019, p. 1934), and describe as less useful than using adult norms. As always, the decision of which corpus to use has to be carefully made considering the aims of the research being undertaken.

With online corpora, test developers can choose or design a corpus or several corpora, ensuring that they are large and representative. Then, a developer may divide this corpus into sections and then into a ranked word list that takes into account range, frequency, and dispersion.

These lists should be checked against other corpora to ensure a thorough coverage of likely known words, and then these lists should be divided into levels based on range, frequency, and dispersion (Schmitt et al., 2020). A researcher may draw on previously established word lists, but they need to ensure that the lists are suitable and provide a strong rationale for choosing those particular lists. For example, there are curated word lists already available (P. Nation, 2016) using data from corpora such as the BNC and the Corpus of Contemporary American English (COCA).

Finally, the sampling procedure for the test needs to be clearly defined (Schmitt et al., 2020, p. 5) and should, ideally, be random yet systematic. A sufficient number of words needs to be selected from each level, and needs to be an amount that is both large enough to ensure good reliability but small enough so that the testing time is not excessively long.

2.2.3. Test Design, Format, and Operationalisation of Word Knowledge

Test format also impacts score interpretation. Common test formats are multiple choice tests, identification tests, and production tests (Bowles & Salthouse, 2008). It is important to consider test format when evaluating scores from vocabulary tests, as different formats cannot necessarily be considered interchangeable. Choosing a test format is linked with how word knowledge is operationalised. Researchers must consider what constitutes knowledge of a word, the different definitions of the words, and the various lexical characteristics surrounding vocabulary (Schmitt et al., 2020). Vocabulary researchers may also need to consider things such as a word's collocations, frequency in different modes, contextual clues, and more.

Most receptive vocabulary tests focus on the form–meaning link, which is the connection between a word's form (whether written or spoken) and its meaning (Aitchison, 2012; Bauer & Nation, 1993; Milton & Fitzpatrick, 2013; Read, 2000). However, what complicates vocabulary testing is that words do not always have a one-to-one correspondence between their form and meaning (Schmitt, 2010). For example, concepts may be represented by just one word or by a multi-

word unit; words in English may also lack formal similarity despite having similar meanings due to their origins from different sources, e.g., Old German, French, Latin (Schmitt, 2010). Things that complicate choosing words include synonymy and polysemy (as an examiner may be testing a form that has several different meanings), literal meanings vs idioms (Kremmel & Schmitt, 2016), and qualities of words such as concreteness or abstractness (Boers et al., 2007; Brysbaert et al., 2014).

Test format is another important point to consider. The test must be designed with the goal of the research and the explicit operationalisation of word knowledge in mind. Two case studies have shown that scores from four test formats (multiple choice, multiple matching, two types of cloze) cannot be taken to mean that words are known any deeper than the form–meaning link, and that developers cannot assume that test-takers can actually use these words when reading, i.e., the test says nothing about word mastery or lexical employability (Kremmel & Schmitt, 2016). This is perfectly acceptable if test designers and researchers are clear on what the test can measure and they are clear on their goals.

Test developers must also carefully assess the problems behind particular item formats and their limitations. Different test formats may lead to a variety test-taking strategies which can result in correct answers despite participants not demonstrating knowledge of these items on the subsequent interview measures. Test-taking strategies can include blind guessing, inferring meaning by considering other word family members, elimination and association, using context-based clues, or inferring from other word forms in the test (Gyllstad et al., 2015; Paul, 1990; Schmitt et al., 2011). Multiple-choice tests (MCTs) are a very popular option, but come with several issues. For example, both guessing and elimination of distractors can mean that test-takers' scores are artificially inflated. The literature has shown that scores on MCTs such as the Vocabulary Size Test (VST) may be increased by amounts such as 16.7% (Stewart & White, 2011), 25% (Stewart, 2014) or even 45% (Kamimoto, 2008) depending on guessing. However, despite these limitations, research has also

found that blind guessing is used as a last resort and that most participants know the meaning of the words, or have partial knowledge and use strategies that indicate this (Gyllstad et al., 2015). Therefore, although guessing is something that test developers must consider, some research has shown that blind guessing is only used when test-takers run out of other strategies and does not, overall, affect test scores. Guessing could even be desirable, as it may demonstrate subconscious knowledge. The important thing is that test developers acknowledge the possibility of guessing, and use an appropriate framework that requires consideration of this and other construct-relevant and construct-irrelevant variables in test design. For example, some tests incorporate and correct for guesswork by the addition of so-called non-words or pseudowords. These words are designed to mimic real words, but are not real words. If participants indicate that they 'know' a pseudoword, it indicates that they are guessing at least some of the time.

Picture-based tests, where the tester or testing program reads and/or presents a word accompanied by a number of pictures, and then the test-taker points or otherwise indicates the matching picture, are particularly suited to children as they have a familiar, simple match-and-point format, are easy to use, and decrease the complexity of vocabulary assessment. However, the use of pictures in a language test has to be carefully considered, as they may be culturally loaded and ambiguous (Bachman, 1990; Bachman & Palmer, 1996).

2.2.4. Existing Vocabulary Tests

So, are there any existing vocabulary tests which meet these standards and could be used to measure the vocabulary sizes of children in the UK? To summarise, a vocabulary test that is suitable for children must be:

1. Not too long, but with enough items to accurately capture receptive vocabulary size;
2. Include a representative sample of low-frequency and high-frequency words, including words which may be more familiar to children than adults.

Several tests will not be mentioned here, including those designed for second language learners, e.g., the Listening Vocabulary Levels Test (LVLT) (McLean et al., 2015), or those which test different aspects of word knowledge other than receptive vocabulary knowledge, e.g., the Word Associates Format, which examines depth rather than breadth (Read, 2004; Schmitt et al., 2011).

Picture-Based Tests

The **Peabody Picture Vocabulary Test** (PPVT) (Dunn & Dunn, 2007), and its British version, the British Picture Vocabulary Scale (BPVS), is perhaps one of the most well-known and most common picture-based vocabulary tests in the fields of education and psychology. It is primarily used in L1 contexts but has also been used in some L2 studies and evaluates receptive vocabulary size, i.e., whether a child understands a word (Groth-Marnat & Wright, 2009). The most current product description available, for the PPVT-4, describes the test as such (Pearson Assessments, n.d.):

> The PPVT–4 scale is a norm-referenced, wide-range instrument for measuring the receptive (hearing) vocabulary of children and adults. Enlarged and colorized, this PPVT edition is available in two parallel forms (Form A and Form B) that are administered individually. Each form contains training items and 228 test items, each consisting of four full-color pictures as response options on a page. For each item, the examiner says a word, and the examinee responds by selecting the picture that best illustrates that word's meaning.

The PPVT can be used to track progress after interventions, predict future performance, or to assess whether there are differences in scores of minority groups (Champion et al., 2003). It is aimed at "individuals with a range of educational backgrounds, including speech-language pathologists, school psychologists, learning disabilities specialists, and educational diagnosticians, who are trained in standardized test administration" (Pearson Education, 2013, p. 1). The PPVT-4 is marketed as a diagnostic assessment alongside several linked interventions, all sold by Pearson Assessments (at the time of writing in 2024, starter test kits with one of the parallel forms are sold at

a starting price of USD$317.50). Software to assist in digital scoring is also offered at subscription prices (starting at USD$50/year).

Thus, despite its extensive use in psychological research, and evidence for its reliability and validity (Bracken & Murray, 1984; Champion et al., 2003; Stockman, 2000; Sullivan et al., 2014), it is clear that the PPVT-4 is not aimed at researchers interested in evaluating vocabulary size directly. The PPVT-4 provides raw and standardised scores, rather than an estimate of vocabulary size, and this out-of-context result does not give us any information about a child's vocabulary (P. Nation & Anthony, 2016). Scores can be compared to standardised figures or another learner's score, and groups of scores can be compared to other groups or norms, so the PPVT enables the comparison of an individual or group to the general population. The test's norms are based on a sample of over 6,000 people representative of the population of the United States (Dunn & Dunn, 2007). However, the PPVT has been criticised for being racially and culturally biased against African American and Hispanic speakers (Stockman, 2000), and has been conceptualised as "a listening technology underpinned by raciolinguistic ideologies" (Cushing, 2023, p. 317).

The **Picture Vocabulary Size Test (PVST)** (P. Nation & Anthony, 2016) is a test of receptive vocabulary size for young first language English speakers up to eight years old. Developed by Paul Nation and programmed by Laurence Anthony, it is based on the most frequent 6,000 word families of English. It is a discrete, selective, context-independent, multiple-choice vocabulary test, using word lists from corpora totalling 5 million tokens. The corpora are suitable for use with young learners and have been checked against Oxford's corpus of children's writing (Banerji et al., 2013). The lists in this test follow a consistent word family definition. The strength of the PVST as compared to the PPVT is that it calculates a child's overall vocabulary size, enabling evaluation of whether a child is above, below, or close to an average vocabulary size for their age. This is not possible with the PPVT because it is not based on a representative sample from word lists. A ceiling effect is

evident with older children as they will have a larger vocabulary size than that shown by the test (I. S. P. Nation & Coxhead, 2021; P. Nation & Anthony, 2016). The test is freely available online (http://www.laurenceanthony.net/software/pvst/ accessed 14 August 2024) and detailed descriptions of the test development procedures are provided in several publications by Paul Nation and colleagues (I. S. P. Nation & Coxhead, 2021; P. Nation & Anthony, 2016). Despite the strengths of the PVST, it has not undergone the same amount of validation studies as the VST, and thus further research is necessary if the PVST is to be used to the same extent as the PPVT.

Pic-Lex is another picture-based receptive vocabulary size test (Alexiou & Milton, 2019) developed for use with younger children. It uses a principled selection of words that learners are likely to know, with a familiar, multiple-choice format, and as its choice of words are simple nouns, it is aimed to be useful for children as young as three or four. One validation study conducted on 40 primary-school-aged children in the UK showed that Pic-Lex measures receptive vocabulary with a fair amount of reliability (Mills & Milton, 2021).

The PPVT, PVST, and Pic-Lex technically measure aural, phonological vocabulary knowledge as the words are read out loud, and in the case of the PVST are accompanied by a short, contextual sentence. This is due to the age of the group being tested, who may be so young that they do not have the reading capabilities required to process a written test. This is an acceptable exchange, but must be noted by the examiners or researchers, as written and phonological knowledge can differ significantly, especially for second language learners (Milton & Hopkins, 2006).

Multiple Choice Tests

The Vocabulary Size Test (VST) is an example of a multiple-choice vocabulary size test which has garnered significant attention, revisions, and research providing validation evidence using both monolingual and bilingual participants (Beglar, 2010; Derrah & Rowe, 2015; Elgort, 2013; Karami, 2012; Nguyen & Nation, 2011). The VST was created using words from the BNC and COCA (Beglar &

Nation, 2007). It measures written receptive vocabulary knowledge and is suitable for measuring the vocabulary size of a native speaker of English. The first version of the VST was based on the first 14,000 word families in English, using a 1 in 100 sample; a newer, 20,000 word version has also been developed, based on a 1 in 200 sample (Coxhead et al., 2015). The former is more suitable for L2 speakers, whereas the latter can be used reliably with L1 speakers of English. There are also several bilingual versions of the VST.

Validation studies of the VST have been criticised (Schmitt et al., 2020) for not investigating potential problems inherent in the original VST, namely examinee test-taking behaviour and the impact of multiple-choice guessing, for which there is no correction. In one study, multiple-choice VST items led to overestimation of vocabulary size by up to 26% at the 9K frequency level in a study on L1 speakers (Gyllstad et al., 2015). A subsequent interview was used to investigate whether test-takers could demonstrate true knowledge of the target words at the meaning–recall level, and thus determine whether they had the necessary knowledge for reading the words on the test. The authors found that approximately 11–26% of items on the VST were correctly answered without knowledge of the target words elicited in the interview format. Furthermore, clear overestimations were observed that were greater than what could be expected from blind guessing (Gyllstad et al., 2015, p. 23). However, they found that guessing had little impact on the scores overall.

Yes/No Tests

Checklist tests, where a candidate only has to answer whether or not they know a word, have been used to measure vocabulary sizes. Examples include X-lex (Meara & Milton, 2003) and the Eurocentres Vocabulary Size Test (Meara, 1990; Meara & Jones, 1989). The former tests knowledge of the first 5,000 words in English and the latter tests the most frequent 10,000 words. Neither are suitable for very young learners, as they both involve reading, but the computer-based formats make them accessible for learners who are able to read and operate the program themselves. Both

tests give an estimate of vocabulary size knowledge. The checklist test format has advantages in its ease of use, and allows learners to draw on aspects of their vocabulary knowledge that other tests do not measure, such as whether they only know the form of a word but not its meaning.

A yes/no test usually includes a proportion of non-words (or pseudowords) to examine whether test-takers are overestimating their knowledge. These words are designed to mimic real words, but are not real words. If they indicate that they 'know' a pseudoword, then their score is adjusted accordingly. Despite the inclusion of non-words, overestimation is still a problem in yes/no tests, as at no point does the test confirm actual knowledge of the word. Whether or not a learner knows the meaning of a vocabulary item is never tested, and as such teachers may find it difficult to accept that yes/no tests are true measures of a learner's vocabulary knowledge (P. Nation & Anthony, 2016).

A checklist test was used in a large-scale crowdsourcing investigation of vocabulary size with success by Brysbaert et al. (2016), who tested 221,268 adults using an online yes/no test composed of 67 items and 33 non-words. A possible future research direction could be developing similar tests for children, maximising sample sizing using computerised methods. However, the feasibility and ethical implications of such testing would have to be considered.

2.2.5. Summary

Methodological problems in early vocabulary size research have led to varied and often inflated vocabulary size measures for English speakers. Newer vocabulary tests have addressed these issues, improving sampling procedures and clarifying things such as the units of count, but some tests are lacking in the critical validation evidence (i.e., empirical evaluation of the test's reliability, validity, as well as documentation and explanation of all aspects of test design) that is necessary if the scores from these tests are to be interpreted accurately and reliably (Schmitt, 2010; Schmitt et al., 2020). The movement towards stronger and clearer justifications in vocabulary size

testing echoes similar sentiments put forward regarding limitations in quantitative corpus linguistic approaches, especially the “tendency for particular features to be operationalised in an ad hoc way, without any clear interest in the interpretive impact of these operationalisations” (Durrant et al., 2021, p. 50). We will return to the following sections of the literature review.

Thus, as the field of vocabulary testing moves forward, test designers must keep several things in mind. Researchers must be clear on their goals for their vocabulary tests; they must describe who the test is designed for and explain the specific purposes for the test. This is important because the goals drive the design process. Then, it is important to design measures transparently, clearly describing and giving justifications for the decisions made in the test development process. The corpus, test format, procedures for choosing a unit of count, sampling and making the word lists should all be described in enough detail that someone else could replicate the process if they needed to. If all of these aspects of test development are considered and addressed, a higher standard of vocabulary test will be produced, and we can more comfortably make inferences and decisions from the vocabulary tests we employ both in practice and in research.

What may help is the development of tests that are freely available on the internet, aimed at supporting the placement of students, identification of issues, or the assessment of progress. In fact, several of the vocabulary tests used in this age are computer-based and thus have the advantage of being easy to administer and score, as well as potentially being available to researchers and practitioners online for free. Feedback is immediately available and test-takers are able to complete the assessment at their own leisure. As well as this, computer-based assessments can access a larger database of items and may randomly select from this database each time, enhancing test security (Alderson, 2001). However, computer-based assessments often rely on a discrete multiple-choice format and are thus restricted to testing one aspect of linguistic knowledge rather than overall communicative skills.

Vocabulary size measures gathered by any research are strengthened with a thorough consideration of test design, validation, ethics, politics, social consequences, and fairness. As language testing is a key mechanism of language policy and control (T. McNamara & Shohamy, 2008; Shohamy, 2001, 2005), it is important to carefully consider not only the technical aspects of designing valid and reliable tests, but also a test's social consequences. If a test score carries with it a meaning that may affect a person's life, then it is not something to be taken light-heartedly.

As a final note, it is important to highlight that this literature review has only considered *receptive* vocabulary size measures and has not considered other ways of measuring lexical knowledge (e.g., productive tests, parental reports for infants, extrapolating from recordings) or the complexity of measuring all aspects of word knowledge (Aitchison, 2012; Milton & Fitzpatrick, 2013). For example, depth of vocabulary knowledge is an important part of 'knowing' a word, including knowledge of its features of phonology, orthography, collocational features, etc. Receptive vocabulary size is only one construct interacting with many other knowledge types (P. Nation, 2013).

2.3 Written Language Development

Writing is a specific communicative language use which enables people to communicate their ideas and emotions through text. Written text has several properties which make it a particularly interesting medium to study. Writing creates a persistent representation of someone's thoughts as they currently exist in their mind, which are then able to be transmitted through both space and time. In contrast to speech, writing can be prepared ahead of time; it can be edited and rewritten, reworked until satisfaction is reached; and the ideas, once written, are fixed in their state. This thesis has been written with a particular communicative purpose and audience in mind, shaping the way in which I have written it. Manipulating language in this manner is complex, involving

several key processes and skills which children develop as they age, and which rely on interacting linguistic, social, cultural, and biological systems (Morrow, 2020; Sulzby & Teale, 1985).

Writing has not been a topic that has historically received as much attention as reading in the literature, although writing research is catching up to reading research (Bazerman et al., 2009, 2017). This section of the literature review will summarise how writing and literacy develop in L1 English speakers. Section 2.4 will introduce some key ideas regarding corpus linguistics and discuss how we can linguistically analyse and quantify written texts, again focusing on writing produced by younger L1 English speakers. For now I will introduce some key concepts related to literacy and writing practices in education.

2.3.1. Early Stages

Learning to read and write marks a significant change in children's development and language experience (K. Nation et al., 2022), introducing more complex words and syntax along with the requisite knowledge of orthography, i.e., the conventions for writing a language, which allow children to link words and sounds with their written representations (Bazerman et al., 2018; Britton, 1972; Morrow, 2020). The term 'writing' can refer to either the process ("I'm *writing* my letter") or the product ("Let's do some class *writing*") (Durrant et al., 2021). An emphasis on the end product of writing (i.e., the things actually written) has, in the past, obscured the idea of writing being a communicative event. Educators or parents may place more emphasis on getting it "right" (e.g., neat handwriting, holding the pencil correctly) rather than simply writing with the aim of communication (Whitehead, 2010).

The acquisition of writing is dependent on the development of oral, phonological, and writing skills (Lonigan et al., 2000; Puranik & Lonigan, 2014; Whitehurst & Lonigan, 1998). Children are most likely to write words they already know and can speak (McKeown et al., 1985; Perfetti & Hart, 2002) and these words are learned from the words they are exposed through conversations

with adults, peers, and siblings, words they are taught at school, and words learned through reading, television, games, and other media sources. Spoken language acquisition always precedes written language, and the assumption that writing is just speech written down has often dictated the research into the act of writing for as long as it has been studied (Harris et al., 2009).

To learn to write, children must understand the idea of representation, i.e. that symbols (in this case, words) relate to particular meanings. Then, copying adults, they produce what can be perceived as random marks, yet communicate that these marks carry a message (Arnold, 1999; Latham, 2002), showing that these marks are not random scribbling (Dunst & Gorman, 2009). In the emergent literacy stage, competence develops through meaningful encounters with print in the processes of early reading and writing (Morrow, 2020; Riley, 1995; Whitehurst & Lonigan, 1998). After play writing, children begin to show *emergent writing*, where they can recognise and replicate some letters (Morrow, 2020). These letters will be chosen randomly at first, then more systematically. This emergent writing shows understanding of the correspondence between written symbols, sound symbols, and meaning. The emergent literacy stage is constrained by physical and neural processes as well as individual differences (Whitehead, 2010). Physical processes include learning to hold a pencil and mastering writing basic letters, while neural processes are related to working memory, phonological and morphological processing skills, oral language skills, and concepts about print (Puranik & Lonigan, 2014). The writing that children produce thus relies on the interplay between all these factors and is influenced by social and cultural contexts. The social and cultural context of literacy development in the United Kingdom will now be examined.

2.3.2. Literacy and Policy

Development of writing beyond this stage is heavily influenced by culture, driven by the complex array of literacy practices that children born into literate societies experience (Lancaster, 2012). It is here that we circle back to Section 2.1, because language policy plays an extremely

important role in the development of literacy in literate regions where school literacy is highly valued (Goodwyn, 2014; Whitehead, 2010). Teaching children to read and write is a key goal of educational institutions in literate societies; children need to learn to read and write in order to progress in educational settings, as writing is a key medium used to transmit and evaluate knowledge. Writing ability is often a key measure of academic success as children progress through the school system (S. A. Crossley & McNamara, 2009) and attempt a number of different written communicative tasks that differ in the skills and resources required. Reflecting this, literacy policies have emerged as key political imperatives globally since the 1990s (Goodwyn, 2014). We will examine some of these policies further in this section, but first a definition of literacy is required.

Literacy can be defined as competence in an area (e.g., digital literacy, computer literacy) or as the ability to read and write. Gee provides the following definition of literacy: "literacy is control of secondary uses of language (i.e., uses of language in secondary discourses)" (Gee, 1987, p. 8). By secondary uses of language, he means using languages in settings outside your 'primary discourse', which is essentially at home, with family (how you learned to use language as a child). An example of a secondary use might be a child communicating with their teacher, for example.

Gee wrote this definition nearly forty years ago, and it is interesting to observe how he describes literacy as multiple types of 'languaging'. This is reflected in the paradigm shift which has occurred in research on literary practices over the last decade, specifically related to how people learn literacy and how to define it. The conventional idea of literacy, one shared by many educators and policymakers (Purcell-Gates, 2020), is that literacy is just the technical skills of reading and writing associated with success in academic settings, i.e., the idea of 'school literacy'. This is the 'old' view of literacy, which is simplistically defined as the ability to read and write, using skills learnt in traditional educational settings. Research has undergone a theoretical shift towards the concepts of multiple literacies (Purcell-Gates, 2020). It is now the dominant view, after scholarly works which

emerged in the 1980s, that you cannot separate literacy from culture, as reading and writing happen in contexts with particular values, and thus literacy always involves social and cultural elements (Heath, 1983; Street, 1984). As Purcell-Gates (2020, p. 4) states:

> From this perspective, literacy development is not seen as linear, building in skill and fluency toward one type of literacy, nor as hierarchical, (e.g., low, functional literacy to high, educated literacy). Rather, it is seen as multiple, occurring across the complex plane of life itself.

Within this view, 'school literacy', or academic literacy, is just but one of many literacies, not just *the* (only) literacy. Academic literacy is shaped by the social, cultural, and political contexts in which it is embedded, involving the academic community and the government deciding what types of literacy are to be valued, taught, and assessed, and is thus "perhaps the clearest example of the ideological nature of all literacies" (Purcell-Gates, 2020, p. 4).

Across educational policy documents in the UK, the themes are similar: language unlocks doors, allows us to communicate with others, and allows access to areas of cultural importance. Language is inextricably linked to literacy, yet separate enough to be distinguished. These documents emphasise that literacy is needed so that children can access the entirety of their education: literacy into a metaphorical key. This metaphor echoes the previous word gap discussion, where language is framed as the bridge that will cross the gap between children who are 'word-poor' and those who are 'word-rich'. These curricula, then, reflect the larger ideology woven through literate societies: that language is good, and more language is better.

Research is beginning to show that children participate in diverse, rich literacy practices outside of the typical, privileged "school" literacies. Whitehead contrasts the idea of school literacy against the idea of literacy being used in its more usual role in shared settings, for example two people interpreting a written document together, or the exchange of emails between colleagues,

which Heath refers to as “literacy events” (1983, p. 200). These shared communicative experiences are often the main experiences of literacy for many people around the world. All of these ways of ‘doing’ literacy and of ‘making meaning’ are valid—literacy is not just what happens in the classroom (Knobel & Lankshear, 2003).

Furthermore, exposure to ‘out-of-school’ literacy practices, as they are often referred to, may have a major influence on children’s development and interpretation of literacy (Knobel & Lankshear, 2003). Research has shown how different cultures and social classes interpret the idea of literacy differently, and how a reliance on traditional ‘schooled’ literacies disadvantages already marginalised groups. These children may be deemed ‘literacy failures’ by the school system but actually participate in their own complex literacy practices that are often deeply embedded in social relationships (Heath, 1983; Wells, 2009). For example, ethnographic research has shown that working-class children may participate in literacy activities at home that are rooted in ‘learning by doing’, and thus find abstract literacy practices in school difficult. Middle-class communities participate in literacy practices that are valued by schools (e.g. knowing fairy tales and writing fantasy stories), whereas working-class and African American communities value other literacy practices (e.g. valuing factual information, independence, participating in wordplay, or inventing rhymes) that lead to confusion when entering the school system (Heath, 1983). These studies show the importance of rethinking what ‘literacy’ means and what skills are valued in the classroom in order to accommodate the complex and varied ways of being literate.

2.3.3. Writing Development in the United Kingdom

When children enter the formal schooling system, the majority of their writing skills, processes and concepts are formalised and practiced. As writers progress through school they must attempt a number of different tasks that differ in the skills and resources required (Guo et al., 2013; Plakans & Gebril, 2013). Progress in this area can be measured over the two axes of time (i.e., age-

related developmental milestones) and quality (i.e., perceived proficiency) (Constantinou, 2019; Durrant et al., 2021). Proficiency in writing relates to communicative competence, the idea that human capacity for language must involve being able to use the required language across different communicative contexts (Hymes, 1972). For children in the United Kingdom, these contexts are primarily dictated by the curriculum of their school. For example, in England, the national curriculum includes an aim for its students to "write clearly, accurately and coherently, adapting their language and style in and for a range of contexts, purposes and audiences", with a specific focus on transcription, composition, vocabulary, grammar, and punctuation (Department for Education, 2014).

In the English national curriculum, literacy development is described under the areas of *transcription* (spelling and handwriting) and *composition* (articulating ideas and structuring them in speech) (Department for Education, 2014). Transcription involves being able to spell accurately through knowledge of both phonology, morphology, and orthography. Writing systems are designed to represent the spoken words of a language and are generally based on two principles: sound and meaning (Steinberg & Sciarini, 2006). English is based on speech sounds, where each symbol in the writing system represents a phoneme or a syllable. Although many writing systems have predictable and corresponding orthographies, where the writing corresponds highly to the spoken word (e.g. Spanish), English does not have a high correspondence between its orthography and pronunciation. This unpredictability makes the English spelling system particularly difficult to learn (Steinberg & Sciarini, 2006).

Once children have mastered writing single words, they then learn to construct written sentences. Sentences in the early stage are simple, using early connectives (*and, then, but*), and gradually become more complex, including main clauses as well as subordinate/dependent clauses. In the early stage, sentence markers (capital letters, full stops) are used haphazardly, and improving

these aspects is the focus of early years education. In England, this is included in the English curriculum for Year 1 (5–6-year-olds) (Department for Education, 2014).

One main aspect that differs between writing and speech is the immediacy of the act. Conversational speech tends not to be pre-planned and once spoken, an utterance cannot be taken back. In contrast, writing can be planned, revised, and edited, and thus the final written text that an external reader might see can be very different to the original text produced. As children age, curricula begin to focus on aspects of text cohesion, mastering complex syntactic structures, and employing cognitive strategies such as planning, editing and revision. At 8–9 years old, there is a marked shift in writing development that reflects "a major cognitive shift and a concomitant growth in vocabulary knowledge required to engage with the accelerating demands of curriculum content in upper elementary school" (Roessingh et al., 2015, p. 68). Planning and revision, as well as understanding of purpose and target audience, are concepts that children come to grips with over their time in education, and are often a focus in educational curricula. For example, the writing curriculum in England mentions 'planning' as early as Year 2 (6–7 years old) (Department for Education, 2014).

Children will now understand that writing is fit for different purposes and be aware of the differences in informative writing versus imaginative writing, for example. Vocabulary becomes more varied, and the interplay between speaking, listening, reading, and writing is prominent, as their English language skills develop as a whole. Writing development past the teenage years places a larger emphasis on planning, organisation, and coherence. The initial planning and drafting stages are more important, and writing becomes organised in systematic ways, with introductions, body paragraphs, and conclusions. There must also be evidence of stylistic adaptations for the audience, conveyed by style, vocabulary, and form (Latham, 2002). The end goal of the curriculum is a writer

who is capable of conveying information in a variety of ways and for a variety of purposes, and, in our society, is capable of using these skills to pass examinations.

2.3.4. Narrative Competence

One of the writing skills children are asked to develop is the ability to write fictional stories and narrative texts. The ability to produce a coherent, fictional piece of writing (a narrative) can be referred to as *narrative competence*. This is the ability to explain events, tell stories and relay experiences, which requires more complex language than normal, daily conversation (Stadler & Ward, 2006). In fact, narration is a key language genre that children must learn to employ in a socially appropriate way, the same as other discourses and registers, as they develop communicative competence (Botting, 2002a). Research shows that children develop narrative competence early in childhood and continue developing it through to adulthood (R. A. Berman, 1995; Montello, 2014) The transition is long because narration is a complex, multifaceted skill that relies on linguistic and cognitive knowledge of genre, structure, and linguistic devices. Children first learn to tell stories orally and then transition while developing their formal literacy skills to expressing written narratives (Mäkinen et al., 2014; Stadler & Ward, 2006). Fiction has particular linguistic features, which be explored further in this literature review, as well as thematic and cultural considerations.

The study of the relationship between oral and written narratives and the development of narrative competence is therefore complicated, because the skills that young children gain in reading, speaking, and writing are interrelated (Lonigan et al., 2000). As oral narratives are the basis for written ones, oral storytelling and development of oral narrative competence have received more attention in the literature than the development of written narrative competence alone (C. Fox, 2003; G. Fox, 2004). In particular, the work by R. A. Berman & Slobin (2013), using the *Frog, where are you?* story (Mayer, 1969), evidenced how different options in language affect children's storytelling abilities. Their research showed that as children age, they show increased coherence and

put more content in their stories (Clark, 2009). It has also been shown that oral narrative competence is associated with literacy ability and that the quality of children's oral narratives predicts reading skills (Reese et al., 2010). Oral narrative competence predicts written narrative competence, particularly children's ability to give structure to their oral narratives; however, oral and written narrative competence do not overlap completely, as cohesion and coherence do not transfer between the mediums (Pinto et al., 2016). Furthermore, oral narratives are often used to assess children with language impairments, as they can show language deficits (Botting, 2002b). For example, narrative ability has been shown to be affected by speech sound disorders (Wellman et al., 2011). Generally, the study of written narratives in L1 English children focuses on specific aspects of vocabulary such as lexical diversity, density, sophistication, or on quantifying syntactical development. These specific quantitative measures will be further discussed in Section 2.4.

On top of vocabulary and syntax skills, during the process of writing development children must learn to manipulate and produce specific features of narrative texts, such as structure, coherence, and cohesion. Furthermore, they must be able to reflect, think, and plan in detail. These abilities are shaped by language that provides children with different grammatical, lexical, and syntactic options (Berman & Slobin, 2013). As children age they also develop the ability to handle and describe more complex temporal states, such as by using complex clauses, passive tenses or conjunctions to connect past and future events. Children learn to add cohesion to their stories by using referential pronouns, and they develop fine-tuned planning strategies. All of these linguistic features aid in more organised, structured stories. These abilities are first acquired orally, then are used and developed in their written narratives, and cemented through formal schooling (Stadler & Ward, 2006).

Out-of-school experiences with literacy also impact written development. Furthermore, motivation is key. Only half the children aged 8 to 18 surveyed by the National Literacy Trust in 2018

stated that they enjoyed writing (C. Clark, 2018). Barriers to enjoying writing included spelling difficulties or trouble deciding what to write. Children acknowledge the functional aspects of writing, agreeing that practicing writing improves their skills, and that developing writing may lead to better job prospects (C. Clark, 2018). All of these considerations have to be incorporated into the classroom.

2.3.5. Summary

Written language development in children is an extremely complex phenomenon that relies on interacting linguistic, social, cultural, and biological systems. The major influence in literate societies is the school system, and, in a wider context, the governmental policies enacted in regard to language education. The role of language policy is key to children's written language development during their formal schooling years. Policymaking impacts curricula as well as examinations and all the related issues involved in language testing. Furthermore, children's writing development is influenced by their out of school literacy practices and a variety of genetic and environmental factors.

2.4 Quantifying Written Language

Sophisticated computer software and developments in corpus linguistics and natural language processing have enabled faster and more reliable illumination into changes and patterns in texts, and large-scale quantitative corpus-based studies show promising advantages for the study of children's writing development (Durrant et al., 2021). This section of the literature review will address how we can quantify children's written language. We begin by discussing the role of corpus linguistics in this sort of research, and then introduce some key metrics used for examining children's writing, focusing on measures to assess vocabulary development (in particular, lexical

diversity, density, and sophistication). We then look at the language used in children's books, and examine how this influences their writing.

2.4.1. Corpus Linguistics

A corpus is a collection of language data selected to represent a language or language variety (Sinclair, 2005). A corpus can consist of spoken or written language; in this thesis we will be focusing on written corpora. Corpora are widely used in linguistic research, especially in studies on child language development (Behrens, 2008). A corpus may be generalised or specialised, depending on the inclusion criteria and research question (Cheng, 2011). Generalised corpora are often very large and allow examination of language as a whole, whereas specialised corpora describe language in a specific variety, register or genre. Specialised corpora are often compared to general corpora to show what language features are under or overrepresented in the specialised corpus (P. Baker, 2010). For example, a specialised corpus, such as a corpus of academic English, could be compared to a generalised corpus, such as the Corpus of Contemporary American English (COCA), to understand which words are more commonly used in academic English compared to 'general' American English.

The study of the development of written L1 English in children has suffered in the past due to practical considerations such as the difficulty in transcribing handwriting (Smith & McEnery, 1998), which has made the collection of large amounts of original data difficult. However, now, corpus linguistic is "inextricably linked to the computer" (Kennedy, 2014, p. 5), as computers allow researchers to create, store, and analyse huge corpora. Researchers now have instant access to huge amounts of data and the tools to allow them to search, sort, count, and accurately show the occurrence of particular linguistic items in a text. Bespoke or ready-made corpus analysis software also allows researchers to search for features in their corpus such as keywords, lexical bundles, frequencies of types and tokens, collocations, and other patterns. Some of these concepts, as well as

some corpus analysis software, will be further explored in the remainder of this literature review, as well as Section 5.

A corpus may never be large enough to include every example of every language feature in a language. A corpus is compiled for a specific reason based on design criteria that ensure the texts are representative of the linguistic feature one wants to study. L1 speaker intuition is still something that has to be relied upon when making judgments about a particular language feature. But what do we mean by representative? And how big does a corpus have to be to be representative? Representativeness can be defined as including all possible categories or groups, but in reality this is difficult to achieve due to the practical problems of defining all the categories or groups. Thus, it is difficult to make one corpus that is truly representative of all people and situations. Larger corpora can attempt to overcome these problems, but how big is necessary?

A corpus can be deemed to be 'big enough' when it contains enough occurrences of the phenomenon under study, i.e., if a researcher is interested in low-frequency vocabulary, then the corpus has to be big enough to include some of these low-frequency words. While much of the prior literature in corpus linguistics has used 'small' corpora (1–5 million words) to look at change/variation, now many favour the use of massive 'mega corpora' because they offer more scope for stronger analyses of lexical and semantic variation (Davies, 2018). However, the caveat is ensuring that the corpus used for studying variation has the appropriate design and structure for measuring variation (i.e., a problem with large corpora composed of billions of words pulled from the internet is that they cannot be subdivided into meaningful sections). Some linguistic features are extremely sensitive to corpus size, such as collocates:

> A one million word corpus is very rarely sufficient for anything but the highest frequency words, and even a 100 million word corpus like the BNC often provides meagre collocates data for moderately frequent words like riddle, nibble, witty, or serenely, which is probably

> not enough to really say much of interest about the meaning and usage of these words. (Davies, 2018, p. 71).

A specialised corpus, even if it is smaller, may provide us with enough occurrences of the linguistic features of interest. This works around some of the problems in representativeness, by allowing us to make corpora which are definitely representative of one particular situation or genre.

Corpus linguistics is continuously developing, and methodological approaches for building corpora are being continuously expanded. For example, the recent development of the *Corpws Cenedlaethol Cymraeg Cyfoes* (National Corpus of Contemporary Welsh) has been thoroughly documented in order to provide guidelines for the development of linguistic corpora in any language (Knight, Loizides, et al., 2021; Knight, Morris, et al., 2021). Furthermore, new studies incorporating technologies such as eye-tracking with corpus research are providing information about the 'process into the product' through combining psycholinguistic and corpus-based data, e.g., in the Provo Corpus (Luke & Christianson, 2018). These are, however, outside the scope of this literature review.

If we want to analyse large amounts of written data, quantitative corpus linguistics approaches seem like the way to go. The following section will now introduce some key concepts in how we can quantify written vocabulary development in children in English.

2.4.2. Linguistic Features of Children's Writing

Quantitative research on writing development in L1 English speakers tends to focus on the areas of vocabulary, syntax, cohesion, and formulaic language. Prior research has prioritised syntactic development but there is now increasing interest in vocabulary (Durrant & Brenchley, 2019a). A systematic literature review identified 104 studies (published between 1945–2015) on language development in the writing of children in Anglophone countries, finding 66 studies on syntax, 45 on vocabulary, 19 on cohesion, and none on formulaic language (Durrant et al., 2021).

This thesis hopes to contribute to this ongoing research into children's vocabulary development in their writing.

Historically, in the field of writing development research, there has been more work carried out with older writers than younger children, normally due to methodological issues in collecting writing samples from younger speakers, and there has also been a stronger focus on writing instruction, writing practices, and multilingualism (Juzwik et al., 2006; McEnery & Wilson, 2003; Smith & McEnery, 1998). Research on writing development is also interdisciplinary, reflecting not just applied linguistics theories and approaches but also educational and psychological research; for example, writing has been examined in terms of its psychological processes in the classroom and as a practice with different functions in different socio-cultural contexts (Klein & Boscolo, 2015).

Research on writing development centres around the core principle of proficiency, which can be defined as "an individual's capacity to draw on their linguistic resources so as to communicate via the written form" (Durrant et al., 2021, p. 17). Proficiency can be evaluated in many ways, and in this thesis we are only going to look at how to evaluate proficiency through linguistic analysis. Three main metrics have been put forward for measuring the lexical qualities of written text: lexical diversity, lexical sophistication, and lexical density (S. Crossley, 2020; Johansson, 2008; Malvern et al., 2004; Read, 2000).

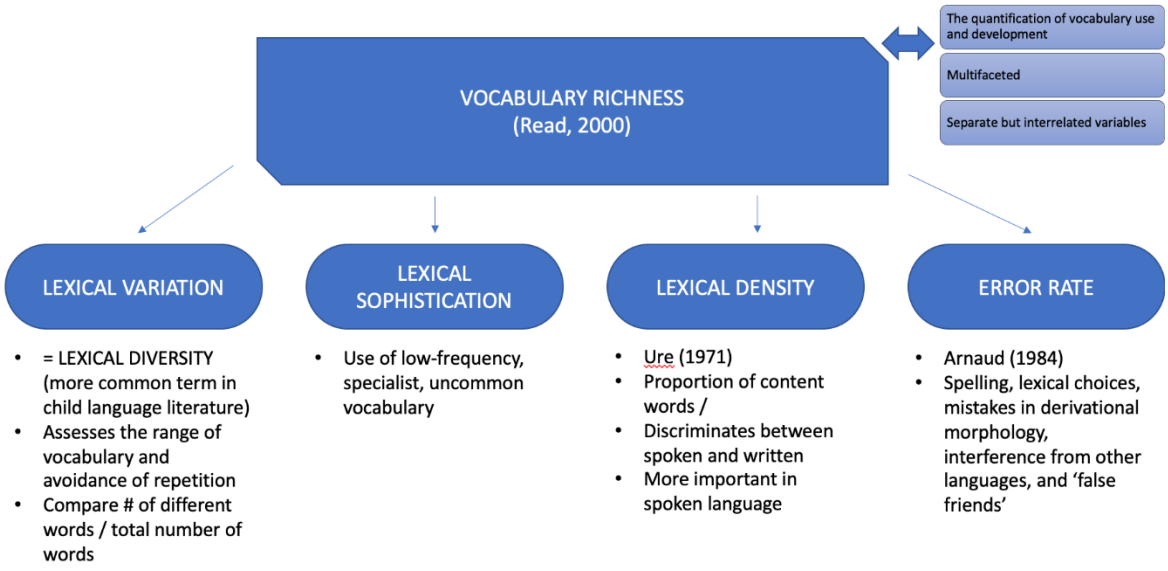
These three ideas can be labelled together by the idea of **vocabulary richness**, sometimes referred to as lexical richness. Figure 2.2 summarises the concept of vocabulary richness, which subsumes the interrelated variables of lexical variation (or diversity), lexical sophistication, lexical density, and error rates (Read, 2000). Lexical richness involves consideration of the range of vocabulary used and how much repetition is in a text, based on the premise that more advanced language use includes not only more words, but less frequent words, with more variety. Notably, these analyses all work under the assumption that "good" writing has particular lexical features

(Read, 2000). In agreement with Read (2000), Malvern et al. acknowledge that lexical richness is "multifaceted and can be measured by interrelated but separate variables" (2004, p. 4).

This chapter will examine lexical diversity, lexical sophistication, and lexical density in turn in order to describe how researchers have attempted to quantify written child language development using these metrics.

Figure 2.2

Vocabulary Richness



Lexical diversity

Lexical diversity is one measure of the multidimensional features of lexical richness (S. Crossley, 2020; Johansson, 2008; Malvern et al., 2004; Read, 2000). Lexical diversity is the assessment of the amount of variation and repetition in a written or spoken text, with the assumptions that a more diverse text has more variation, less repetition, and so signifies a higher quality of language employment. This section will first present the development of *methods* in

measuring lexical diversity, and will then look at the results of lexical diversity research in children's written language.

Diversity is predicated on the separation between types (the number of distinct words) and tokens (all words in a text), and the simplest methods for measuring diversity began by investigating these type–token ratios. The most basic measure of lexical diversity is the number of different words (NDW). However, a methodological problem with NDW is that samples of different lengths cannot be compared because as texts increase in length, the number of different words also increases. Even though studies have used transcripts standardised by length, there is no guarantee that the language samples of the individual will all be of the same length, and length affects results (Malvern et al., 2004, p. 17). This led to the creation of the *type–token ratio* (TTR), which relates the number of different words (types) to the total number of words (tokens). A high figure means that the text contains many different words and ergo that the writer/speaker has a larger, more varied vocabulary; a low figure means that the text does not contain as many different words and words are repeated. In theory, the TTR seems like a useful measure of vocabulary diversity, and is easy to calculate. It has been used in many research studies as both an input variable and a child language variable (B. Richards, 1987). However, the TTR measure, and subsequent measures that aim to improve on TTR, have several issues related to (i) size of the text, (ii) proficiency of the learner, and (iii) lexical frequency.

The size of a text will affect LD because a larger sample will have a smaller ratio of types to tokens. Eventually, TTR ends up as a function of sample size, and not a true representation of lexical diversity. It is imperative that any studies using TTR also control for the size of the sample (McKee et al., 2000a). If size is not controlled, then as the size of the sample increases, the availability of new words decreases, and thus the TTR inevitably decreases.

Essentially, with each new token, there is a corresponding decrease in the likelihood of a new type (McCarthy & Jarvis, 2010, p. 382) because some repetition has to occur in any text. So, longer texts naturally give decreased values of diversity, even if they may be complicated or highly diverse texts. All texts also eventually reach a point of "saturation", where all the necessary types have been used, and no new types are being encountered. When this happens the text or corpus can be said to be fully representative of the word types that are indicative of that text's theme (McEnery & Wilson, 2003). A further problem related to text type and size is its subsequent homogeneity, i.e., different types of texts will naturally have different diversity levels due to their rhetorical purposes and strategies and so text type/genre must be suitably considered when considering lexical diversity.

A learner's proficiency can also impact lexical diversity scores. For example, as proficiency increases, types and tokens also increase as learners employ larger and more varied vocabulary. If there is a linear relationship between types and tokens, then no development will be able to be measured by TTR because it will always have a constant value (van Hout & Vermeer, 2007). Different stages of proficiency will also lead to irregular patterns in measuring TTR. An example given by van Hout and Vermeer (2007) is the learning of function words such as *an* and *the*, which are very frequent words that will increase the amount of tokens a learner uses. However, the number of types will only have increased by two and the TTR will decrease, despite the student becoming more proficient. Therefore, it is important to distinguish between content and function words.

Thirdly, frequency of words affects the TTR; taking frequency into account will give better measures of lexical richness by relating the words in a dataset to their difficulty, which is related to lexical sophistication (Van Hout & Vermeer, 2007; Vermeer, 2001).

Scholars have attempted to solve these problems, notably the problem of text length, in several ways through mathematical methods. One measure is the index of Guiraud (G), which uses the square root for relating types and tokens (G = types / √ tokens). Another metric, Maas, is a log

correction index which shows no effect of text length if analyses are limited to certain ranges. However, validation evidence is lacking for Maas and other log correction indices (McCarthy & Jarvis, 2007).

One of the more successful solutions takes into account the number of low-frequency words, as low-frequency words occupy the largest percentage of words that people use on a day-to-day basis, through calculating a parameter called D. Proposed by David Malvern and Brian Richards, D calculates the probability of a new vocabulary item being introduced into longer samples of the text by finding the best fit between empirical and theoretical curves of TTR versus tokens (Malvern et al., 2004). The calculation is the result of a series of random text samplings by taking 100 random samples of 35 tokens, calculating the TTR for each sample, then storing the mean TTR. This procedure is repeated for samples from 36 to 50 tokens, and then a TTR curve is created from the means of every sample. Then, a formula is carried out that calculates a theoretical curve that most closely fits the empirical curve formed from the random sampling procedure.

The value that creates the best-fitting curve is D and this value usually ranges from between 10 to 100, with higher values indicating greater LD. The procedure is run three times, and the average D value is the final output. According to McKee et al. (2000b), D is a valid and reliable measure of linguistic diversity. It overcomes problems surrounding the previous use of TTR, and has the following stated advantages: "(1) it is not a function of the number of words in the sample; (2) it uses all the data available; (3) it is more informative because, as opposed to a single value of TTR, it represents how the TTR varies over a range of token size for each speaker or writer" (McKee et al., 2000b, p. 324). Early testing using CHILDES provided evidence for the reliability and validity of D and found that if they used random sampling, the results of D were not a function of sample size (Malvern et al., 2004). Random sampling is thus integrated into software for calculating D as standard.

However, D has been criticised. Its notable shortcomings include evidence of significant variation in text length, despite assertions to the contrary (McCarthy & Jarvis, 2007), as well as the suggestion that D merely replicates another measure, namely HD-D (the hypergeometric distribution diversity index), and is thus just "an approximation of a well-established probability function" (McCarthy & Jarvis, 2010, p. 382). HD-D is an approximation of the hypergeometric distribution, which is the probability of drawing a certain number of tokens of a particular type from a sample of a particular size (McCarthy & Jarvis, 2010). Further criticisms of D include that it performs unsatisfactorily in concurrent validations and still tells us little about the quality of the lexicon in measured texts (Van Hout & Vermeer, 2007). D may also not produce results that correlate with other measures of language proficiency or that discriminate well between groups with obvious differences in vocabulary. For example, in one study Finnish and Swedish schoolchildren had higher scores for D than native English speakers of the same age, suggesting some influence of L1 on D (Jarvis, 2002).

Another measure of lexical diversity is the measure of textual lexical diversity (MTLD) (McCarthy & Jarvis, 2010). This is a measure of diversity based on the average number of tokens it takes to reach a given TTR value. MTLD is calculated by sequentially evaluating each word of a text and assessing whether the word is a new word. This happens until a default TTR factor size value (.720) is reached. Once this is reached, the TTR evaluations are reset and the factor count increases by 1, then the process occurs again in the next word of the text. At the end, the total number of words in the text are divided by the total factor count to produce the final MTLD value, i.e., the average of all the factor lengths. The texts are processed both forwards and backwards and the mean of these two values is the final MTLD value. The MTLD value represents the average number of words needed before a text reaches a point of *stabilisation* where "neither the introduction of repeated types nor a considerable string of new types can markedly affect the TTR trajectory"

(McCarthy & Jarvis, 2010, p. 386). In other words, MTLD is “an estimation of the point at which TTR stabilizes in longer L1 texts” (Zenker & Kyle, 2021, p. 5). Several studies have indicated that MTLD is a valid measure of diversity related to writing quality (S. A. Crossley & McNamara, 2009; D. S. McNamara et al., 2010; Zenker & Kyle, 2021). MTLD has been shown to perform well in an initial validity study where it correlated with other measures of lexical diversity (D, Maas, and HD-D), and it does not seem to be affected by text length (McCarthy & Jarvis, 2010). Zenker and Kyle (2021) corroborated this research, summarising that the MTLD Original measure remained relatively stable across texts in the range of 50 to 200 tokens.

To conclude, the best method for assessing lexical diversity is not to rely on one measure, but to incorporate a variety of different yet complementary measures. This is because LD can be assessed in many different ways using indices that capture “unique lexical information” (McCarthy & Jarvis, 2010, p. 391). A combination of HD-D and MTLD will provide a researcher with different information on the construct under investigation. Van Hout and Vermeer support this suggestion by recommending the combination of three elements: consideration of lexical frequency; the use of alternative functions of the type–token ratio (e.g., D and MTLD); and resampling procedures (Van Hout & Vermeer, 2007, p. 122). Whichever method is chosen, a researcher investigating LD has to be aware of the influence of text length otherwise they risk unknowingly confounding their results. They must also consider text type and genre, as different texts will have different diversity levels. Put simply, vocabulary must be assessed in a way that relates to the context in which the text has been written (McCarthy & Jarvis, 2010).

The research investigating lexical diversity has suggested that it significantly correlates with age, i.e., lexical diversity increases as children get older (R. A. Berman & Nir-Sagiv, 2004; Chipere et al., 2001; S. A. Crossley et al., 2011; Durrant et al., 2021; Johansson, 2008; Malvern et al., 2004; Olinghouse & Wilson, 2013; Uccelli et al., 2013). When testing D (one measure of lexical diversity)

using samples of early child language taken from the Bristol Corpus, Malvern et al. (2004) found a significant increase in D with increasing age, with consistent upward trajectories and common patterns in the data with young children (aged 18–60 months). After standardising the transcripts, the researchers evaluated the mean TTR and D scores, and found that TTR decreased with age while D increased with age. This showed a significant correlation between D and age. For 5-year-old children, a mean score of 64 was found (range, 5–83). This increase in D with age has been found in other studies in various populations. In the American context, Crossley et al. (2011) assessed the writing development of older students and found that D increased with age from Grade 9, to Grade 11, to college-aged young adults. Berman and Nir (2010) also found a significant increase in D with age, especially with writing samples when compared to speech.

In a study with UK schoolchildren, Chipere et al. (2001) analysed 899 narrative essays from children of 8 to 15 years old. Scripts were categorised according to Key Stage, and within in each Key Stage they were graded according to level of writing ability. *vocd* software was used to extract values for D for all essays. Comparing D values with key stage and level, the authors found that LD was related to age and writing ability. However, the mean lexical diversity scores were higher for KS2 children in comparison with KS1 and KS3, although not to a statistically significant level. The authors conclude that KS2 scores may possibly be higher due to expected dips in performance in KS3; or, older pupils may produce more coherent writing due to using a greater level of lexical repetition, which leads to lower diversity scores. This research raises several questions, such as whether increasing knowledge of morphological derivations leads to increased lexical diversity. If morphology does impact lexical diversity, then this has theoretical and pedagogical implications as it suggests that morphological knowledge grows gradually during the school years but at different rates for every student (O'Donnell, 1976).

The genre of text also significantly impacts diversity scores. D values for persuasive essays written by high school students were on average 72.25 (Uccelli et al., 2013), compared to values reported for high school students' narrative texts (70–80) and expository texts (80–90) (Berman & Verhoeven, 2002). Olinghouse and Wilson (2013) used MTLD, another measure of lexical diversity, in a study evaluating vocabulary across different genres, and found that narrative texts presented higher MTLD values than informative or persuasive texts, and that MTLD correlated positively with quality ratings for narrative and persuasive writing, but not informative. These values show that different genres of writing draw on different linguistic and rhetorical features, and highlight the importance of carefully considering and reporting on the genre when discussing any measurement of lexical knowledge.

Lexical sophistication

Lexical sophistication involves quantifying the frequency and the relative difficulty of words employed in writing. So how do we decide which words are more "difficult", when words may be used in different contexts by different speakers (Meara & Bell, 2001), and what may be difficult to one speaker may come second nature to another?

Many assessments of writing proficiency incorporate frequency-based measurements (see Section 2.2.2 for a prior discussion on frequency and language acquisition). Less frequent words are considered more sophisticated (Kyle & Crossley, 2015) and researchers assume that those with higher proficiencies and larger vocabularies are more likely to use less frequent words (Meara & Bell, 2001). Research supports the idea that "rare words" indicate higher language proficiency (Arnaud, 1984; Linnarud, 1986). In general, research on lexical sophistication has shown that it is a predictor of age, with scores able to distinguish high school and college student essays (Freedman & Pringle, 1980).

However the research is not as conclusive as the research on lexical diversity, mainly due to how sophistication is operationalised (Durrant & Brenchley, 2019b). The definitions of difficult, rare, or sophisticated words varies (Meara & Bell, 2001). Read (2000) defines lexical sophistication as the use of low-frequency words that are outside of 'general' vocabulary. Researchers can operationalise sophistication by counting the proportion of words in the analysed texts that are not in high-frequency vocabulary lists. Frequency-based measures are always reliant on the definition of a word and the frequency lists chosen. Recent research (Brysbaert et al., 2018) has highlighted the need for more robust and modern word frequency lists, such as those based on subtitles (e.g., SUBTLEX-UK, van Heuven et al., 2014). There are a number of tools that have been developed to measure lexical sophistication, including Range (Heatley, Nation & Coxhead, 1996), P_Lex (Meara & Bell, 2001), The Tool for the Automatic Analysis of Lexical Sophistication (Kyle et al., 2018), Coh-Metrix and VocabProfile (Cobb, 2013). These tools will be discussed in the following section.

As children age, they employ more low-frequency vocabulary (Olinghouse & Leird, 2009; Olinghouse & Wilson, 2013; Sun et al., 2010). These three studies, analysing texts collected from young American schoolchildren (ages 7–11) converge on the same result through different methods of operationalising sophistication. Olinghouse and Leird (2009) compared written texts to the 'Basic Vocabulary Spelling List', a list of the most common 850 words students use when they write. Words used by the children in their study were considered infrequent if they did not appear on this list (Olinghouse & Leird, 2009). Olinghouse and Wilson (2013) used a similar methodology but compared the children's writing to the General Service List (West, 1953). Y. Sun et al. (2010) used a program called Range (Heatley et al., 1996), which can determine the proportion of words in different frequency bands. Using this, a higher percentage of words in the first 1,000 most frequent words was found to be associated with younger children, with higher amounts of words from the later bands associated with increasing age (Y. Sun et al., 2010). This demonstrates how as children age,

they draw on an increasingly varied vocabulary and employ more low-frequency words in their writing.

Coh-Metrix is a tool which measures frequency using a reference corpus (CELEX) in order to calculate a frequency count for each word in a text, which can then be compared to the texts inputted into Coh-Metrix for evaluation. Using Coh-Metrix, Crossley et al. (2011) found that several variables distinguished between grade levels and one of the key ones was word frequency. S. A. Crossley et al. (2011) found that ninth grade (ages 14–15) essays showed significant differences compared to college freshman (age 18) essays in terms of word frequency values, but there was no difference between ninth and eleventh grade essays; in other terms, the ninth-grade essays contained more high-frequency words and the college essays contained more low-frequency words. This suggests that there is some development in vocabulary sophistication between ninth grade and college, but due to the lack of analysis for other school years, it is hard to fully compare trends in this development.

Sophistication can also be linked to the perceived quality of a text. Increasing amounts of low-frequency words align with increases in text quality (Malvern et al., 2004). Roessingh, Elgie and Kover (2015) studied the writing of 8–9-year-olds in Canada using VocabProfile (Cobb, 2002). VocabProfile can be used to compare writing samples to frequency lists. In total, they analysed the writing of 77 children who produced pieces of persuasive text during an in-school session lasting 45 minutes. Their aim was to make quantitative insights into the written vocabulary development of this cohort of children, and attempt to improve upon the methodological problems of prior studies that had limitations on computer tools and lack of motivational prompts or time. They found that length of text, number of words, and percentage of “off-list” words correlated with subjective ratings of the writing samples. Furthermore, they concluded that better writers not only produce longer texts but also use more of their available vocabulary.

Frequency of *register-specific* vocabulary can also be another metric for measuring lexical sophistication, such as the proportion of "academic" words used in a text. For example, studies have counted the number of words in texts written by children that are on the Academic Word List (Coxhead, 2000). The capacity to evaluate this is included in software such as Range and VocabProfile. Results are mixed: while Sun et al. (2010) found a significant increase in the proportion of words from the AWL as children aged, Olinghouse and Wilson (2013) excluded this variable from their study due to low levels of occurrence. Other studies have looked at children's use of Greek- and Latin-based words, which have shown to increase with age (Berman & Nir-Sagiv, 2007; Berman & Nir, 2010). This indicates that as children get older, they move towards using a greater proportion of academic vocabulary. Further investigations into the impact of academic vocabulary on lexical sophistication measures are necessary.

Recent research shows that the concept of sophistication is not as straightforward as it initially seems, as frequency interacts with parts-of-speech, lexical diversity, and register. Durrant & Brenchley (2019b) created a bespoke corpus totalling 2024 literary and non-literary texts written by British children in Years 2, 6, 9 and 11 (ages 6–7, 10–11, 13–14 and 15–16, respectively) in order to evaluate lexical sophistication. They modelled sophistication as low-frequency, register-appropriate words (Read, 2000), but found that this model may not be the most appropriate due to discrepancies found in the data. For example, younger children's writing was characterised by repetition of high-frequency verbs and adjectives but low-frequency nouns. This was due to the use of nouns that do not occur in general adult-directed writing, e.g. *fairy*. They conclude that lexical sophistication cannot be separated from lexical diversity, as is often done in writing research, but rather that they interact with each other; furthermore, other factors such as collocations or syntactic variables may be worth further investigation. This study highlights the importance of measuring a range of variables in order to view a complete picture of children's writing. It also shows how

something as straightforward as frequency needs careful consideration and operationalisation to ensure that the gathered results are meaningful. This is especially important in lexical analyses, as frequency has a big effect on vocabulary acquisition and use.

Longer words are often considered to be more difficult, and so word length has been assessed as a component of lexical sophistication. Word length is a variable that impacts lexical processing time, with research showing a U-shaped curve for reaction times based on word length (i.e., longer reaction times for both shorter and longer words) (New et al., 2006). Research has assessed the number of letters per word, finding no significant age effect (Malvern et al., 2004) but some relationship with quality (Malvern et al., 2004; Myhill, 2013). The number of syllables per word has been shown to increase in line with age and quality. For example, older writers will use more words with more than three syllables and these longer words have some influence on increased quality (Olinghouse & Leaird, 2009). Despite these findings, word length is usually not regarded as important in lexical sophistication measures as frequency.

Lexical density

Lexical density most often refers to the amount of "content words" in a text, i.e., nouns, verbs, adjectives, and often adverbs, in proportion to the total amount of words. Content words provide us with an understanding of the amount of information that a text contains, i.e., a text with more content words rather than function words (prepositions, interjections, etc) contains more information. Density can be calculated in a number of ways, such as by simply dividing the number of content words by the total number of words in a text (Johansson, 2009; Berman & Nir, 2010; Hall-Mills & Apel, 2015) or by determining the number of content words per clause (Uccelli et al., 2013). Density is often included in measures of lexical richness but presents mixed results. Some studies on lexical density in written texts have found no significant differences across age groups (R. Berman & Nir, 2010; Hall-Mills & Apel, 2015) or correlations with quality (Uccelli et al., 2013). In a study on

developmental aspects of text production in both writing and speech, Johansson (2008) attempted to evaluate the developmental patterns in lexical diversity and density in a sample of older Swedish schoolchildren, in order to determine whether lexical diversity or density are sensitive to genre and modality. The results showed no differences between 10–13-year-olds and 13–17-year-olds for all text types, showing that if there is an age factor involved in an increasing lexicon, it will not be obvious unless long-term development is looked out. There were bigger trends found for diversity measures than for density measures, so diversity is a better aspect to look at if one wants to measure differences between age groups. Other research has shown slight increases in lexical density with age, but no significant differences (Johansson, 2009). Density may be sensitive to genre, with narrative texts showing a higher density compared to expository texts; furthermore, written texts showed a higher density than spoken texts (Johansson, 2009). Overall, compared to lexical sophistication and diversity, lexical density appears to less important in relation to development (Durrant & Brenchley, 2019).

2.4.3. Linguistic Features of Children's Literature

Here I will quickly present some of the research which has been conducted on children's literature, i.e., on fiction books written for children by adults. Although this thesis is not examining children's literature itself, it is important to consider children's books as one source of language input in their development. Linguistic features are heavily dependent on genre and register (Biber & Conrad, 2019) and it is important to note that children's literature represents not just one genre, but several interrelated 'strands' or types of text, that may overlap linguistically with other types (Knowles & Malmkjær, 2002).

Children's literature can be analysed using the same methods as described in the previous section. However, corpus-based approaches to analysing children's fiction are rare: "The language of

children's literature receives little explicit attention" (Stephens, 2006, p. 73). More focus has been placed on the themes and content of children's books, rather than the linguistic features being used.

Before we examine some of these specific linguistic features, it is important to highlight that the language of children's books can be considered a powerful socialising instrument (Halliday, 1978) through which a child learns customs and attitudes. How children's literature is used as a vehicle for adult ideologies has been "one of the most fertile lines of inquiry in children's literature scholarship since the field's origins" (Ford Smith, 2021, p. 1). The literature that children consume at home and at school (either on their own, or because of caregivers/teachers) is a major influence on their cultural development in general and influences their writing as they age (Hunt, 2006; McCallum & Stephens, 2011; Sarland, 1998). Children's books represent a significantly unbalanced power relationship, as children's literature is aimed at children but written by adults, whose social and economic power exceeds that of children (Ford Smith, 2021; Knowles & Malmkjær, 2002). As adults adapt their writing to suit children, they rely on certain assumptions about the reader, and, unknowingly or not, showcase their own ideologies.

Modern children's literature has its roots in the mid-nineteenth century, as publishers realised there was a significant market for children's books, and the genre has evolved since then alongside changing social values and issues surrounding gender, race, and more, to what is nowadays a varied and prolific area of literature (Knowles & Malmkjær, 2002). As children learn to read, and immerse themselves in the typical genres of children's fiction, such as fairy tales, ghost stories, fantasy, and, as they age, teen romance or coming-of-age novels, children are exposed to distinctive writing styles and learn what constitutes a 'good' story. In the nineteenth century stories tended to focus on adventures, schools, and, later, fantasy and imagination. Gender held a central role in early children's literature, marked by the idea of stories 'for boys' or stories 'for girls'. Characters were one-dimensional and there was a heavy emphasis on morality. The early twentieth

century produced little that was original, but post-WWII magic and mythology became central themes of children's books, and now they tend to contain both exciting narratives as well as moral truths or lessons (Knowles & Malmkjær, 2002). An increased emphasis on psychological and ethical issues, portrayed through complex characters, inevitably reflects writers' preconceived notions about good and evil (Nikolajeva, 2004). Knowles and Malmkjær (2002, p. 32) highlight what they call *institutions* that are prominent in children's fiction: family, friendship, gender, home, race, and religion. Through writing about these institutions, authors 'locate' the reader while representing their own world view.

Another feature of children's literature is the presence of a strong narrative voice, i.e., point of view. The narrative voice of a story allows interpretation of the scenes for the reader, and controls how a reader understands the text (Stephens, 2006). Readers tend to align themselves with a first-person point of view or with a focalising character, with focalisation being "selection or restriction of narrative information in relation to the experience and knowledge of the narrator" (Niederhoff, 2011, p. 1). Children's literature has a "moderately sophisticated use of focalisation", according to Stephens (2006, p. 84), as well as a high proportion of conversation, direct speech, and communicative events.

In contrast with informational, noun-heavy texts, fiction tends to contain a lot of action and descriptions of action, so include more verbs and adverbs in order to describe actions and events (Biber, 2019). Narrative texts often capture aspects of oral language, using familiar language and simpler syntactic constructions, with shorter, high-frequency words (Biber, 1988). Neate (1999) lists the main differences between informational (or expository) and narrative texts: expository texts are more concise, with technical vocabulary; expository texts use present tense rather than the past; expository texts use the copular and relational verbs rather than action verbs more often used in narrative texts; there are fewer pronouns in expository texts; there are more words between the

subject and the verb; and fewer connectives are used in expository writing. In a study on the vocabulary content of children's picture books, Montag et al. (2015) found that picture books have a greater lexical diversity and density than child-directed speech. These 'book words' are more commonly nouns and adjectives, tend to be longer and more morphologically complex, are more abstract, and are more emotionally arousing (Dawson et al., 2021; K. Nation et al., 2022).

In one comparison between a corpus of fiction written for adults and a corpus of fiction written for children (all from the BNC), Thompson and Sealey (2007) found remarkable similarities between the two in terms of frequency data for words and parts of speech. This suggests core vocabulary consistent with narrative texts, a lexicon that differs when it was compared to a news corpus. Furthermore, they found that children's texts have a large focus on living creatures, objects, plants, food, communication (i.e., direct speech), with an emphasis on size, speed, bravery and fear (Thompson & Sealey, 2007, p. 15). This aligns with findings from an analysis of the Oxford Children's Corpus of fiction written for children which found a greater emphasis on the natural world and physical space (Wild et al., 2013). Adult fiction, on the other hand, when analysed from a corpus linguistic perspective, is more concerned with time, intimacy, law and order, society, and beliefs (Thompson & Sealey, 2007; Wild et al., 2013).

Upon further analysis of n-grams, Thompson and Sealey (2007) found some subtle differences between a number of common words such as *finger*, *neck*, and *angry*, suggesting that children and adults will employ core vocabulary in different ways. The corpus used in Thompson and Sealey (2007) has been compared to a corpus of writing by Dr Seuss (Foster & Mackie, 2013), and a fair amount of overlap was found between the two in regard to lexical verbs and adjectives, although they appeared in much lower frequencies. There was not a lot of coverage for nouns, mainly due to the specific vocabulary employed and repeated in works by Dr Seuss (e.g. *cat*, *fish*, *zoo*).

2.4.4. Conclusions

Quantitative analyses of children's writing have huge practical implications for measuring ability, discovering changes over time, and mitigating issues such as the word gap. Yet, despite increased interest in the vocabulary use in L1 English writing, there is a gap in the literature in relation to a large-scale analysis of children's writing. This gap is especially pronounced for the analysis of children's creative or narrative texts, as the sorts of texts analysed in linguistic research are often expository essays.

Whether it's writing a text or jotting down a shopping list, writing is a skill that adults give little thought to. However, writing is a complex phenomenon and learning to write is a gradual process that relies on a child drawing on a number of complicated cognitive and physical processes. Writing is assessed in academic contexts as a child proceeds through school, with various expectations placed upon the child. Children are encouraged to write in different genres (expository, narrative, persuasive) with emphasis placed on the elements of writing that practitioners deem the best. For example, a key expectation of successful writing nowadays is "control over a large, precise vocabulary" (Roessingh et al., 2015, p. 68).

This degree of control over a lexicon, the ability to employ a wide variety of words, is often referred to as lexical diversity or lexical richness. Measures to measure lexical richness have varied through the years, with frequency-based measures, type–token ratio, and *D* all having been discussed in this literature review. Research has shown that "better" writers (i.e., those who score higher on trait-based rubrics) have higher scores of lexical diversity (Chipere et al., 2001; Roessingh et al., 2015); furthermore, lexical diversity measures often correlate with age (Malvern et al., 2004).

This literature review reflects the wide variety of methodological approaches to the study of children's writing. While some authors have focused on syntactic complexity, analysing metrics such as T-units, others have focused on evaluating diversity using D, or have implemented frequency-

based tools to quantify coverage of high-, medium- and low-frequency words. These variables have been compared against subjective ratings or evaluations to evaluate correlations with quality, or development has been tracked over time to evaluate correlations with age. The results show that as children age, lexical diversity increases (Berman & Nir, 2010; Crossley et al., 2011; Malvern et al., 2004; Olinghouse & Wilson, 2013; Uccelli et al., 2013); lexical density does not show as strong of a relationship; and lexical sophistication presents mixed conclusions, as it may interact with other factors such as repetition, frequency, collocations, or otherwise (Durrant & Brenchley, 2019). More work is needed in these areas to uncover the connections between different lexical measures. Furthermore, there is a dearth of studies on formulaic language in L1 writing development (Durrant et al., 2021).

Another important thing to note is that the authors of prior studies often highlight the variability of writing samples received, which are samples generally collected in controlled conditions to form small, bespoke corpora. The analysis of larger corpora of texts will allow for deeper investigation into the lexical features of children's writing and provide robust evidence for the study of the development of children's L1 writing.

3. Vocabulary Size and Growth in a Sample of British Primary School-Age Children

Vocabulary size and growth in children is a topic of interest for those investigating language acquisition and its relationship with education. Understanding vocabulary size and growth is important in designing practical vocabulary programmes based on what students should know, already know, and need to be taught (Graves, 1986). However, the assessment of vocabulary size and growth is notoriously tricky and requires careful consideration and operationalisation of several key concepts (e.g., what is a word, how can we test knowledge of words, productive vs. receptive vocabulary, etc.). The preceding literature review discussed the background to vocabulary size testing and presented some examples of vocabulary size tests which can be used with L1 English-speaking children. The following literature review will present a discussion of some of the results of investigations into children's vocabulary sizes.

There is a great amount of variation both in the estimated vocabulary sizes of children and the rate of growth (i.e., how many new words are learned each year). Research on vocabulary sizes in children and adults has been ongoing since the 19th century, but methodological shortcomings limit these results (Graves, 1986; Lorge & Chall, 1963; T. G. White et al., 1990). These shortcomings include overestimations of vocabulary size due to how 'word' is defined, as well as the difficulty in designing accurate sampling procedures. Measurements of vocabulary size now sample from substantial lists of English words, a process made much easier with the help of computers (I. S. P. Nation & Coxhead, 2021).

It has been estimated that preschool children know around 100 to 200 words by the age of 2. These words are at first acquired slowly, over several months, and then vocabulary begins to accumulate at a faster pace, with a period of rapid vocabulary learning (the vocabulary spurt) occurring around the time children have acquired 50 words (Goldfield & Reznick, 1990). Vocabulary size testing of very young children relies on parental vocabulary reports such as the Communicative

Development Index (Fenson, 2007) because infants cannot be expected to complete vocabulary testing on their own. Diary methods and home-based recordings provide an estimate of what a child knows but may lead to a sub-sampling of the real lexicon if recordings are short, while longer recordings can suffer from masking new acquisitions (Mayor & Plunkett, 2009). These diary-based methods thus gather productive vocabulary size estimates, based on parents remembering all the words their children have produced. The difficulty is recording enough data to gain an accurate estimate, as young native speakers' vocabulary grows very quickly, outstripping the ability of a human recorder to keep up (P. Nation and Anthony, 2016).

As children age, receptive vocabulary tests can be used, whether these are bespoke or previously created tests. Many early receptive vocabulary size tests used dictionary-based methods, which led to problems with representative sampling. Diller (1978), using dictionary-based methods, put forward that high-school students learn 20,000 words a year and know 50,000 words by the age of 7 and 100,000 by high school (age 16). However, these figures do not seem feasible as they would involve learning over 50–60 words a day. This reflects the difficulties encountered when using dictionary-based methods where a dictionary is said to represent the word stock of a language and children are tested based on a 'representative' sample of those words (Goulden et al., 1990; Nagy & Herman, 1987).

To correct for these methodological shortcomings, Nagy and Herman (1987) recalibrated five older vocabulary size studies to correct for the estimated total number of words in English (as previous studies had underestimated the total word stock of English). Their recalibration suggested that children will know around 10,000 words by the age of 8–9. However, these results vary hugely, with figures ranging from 4,000 to 24,000 words. Another study (T. G. White et al., 1990) found that by the age of 9–10 years old, the children sampled were estimated to know between 12,000 and 16,000 words depending on the socioeconomic background of the school where the children were

evaluated, with the children from the school in a high-SES area learning more words than the children from the school in a low-SES area.

Anglin (1993) estimated 6-year-olds would know around 10,000 words and 10-year-olds could recognise an estimated 40,000 words, suggesting a growth rate of roughly 20 words per day. An important caveat of this extensive research is that Anglin highlights that many of the words that children were estimated to 'know' could actually have been understood through morphological analysis and decoding. Taking away these types of words, Anglin suggests that the children would know between 6,000 and 20,000 root words from the ages of 6 to 10 (halving the daily growth rate).

More recent estimates have led to a suggestion of around 2,000 to 3,000 word families learned per year (P. Nation, 2013). Nation's rule of thumb is that a vocabulary size estimate for a native speaker can be calculated by taking 2 years away from the child's age and multiplying the result by 1,000 ((age – 2) * 1,000). The vocabulary size scores of particular learners may be 1,000 to 2,000 words on either side of this average and even the lowest scoring native speakers at most levels know several thousand words (I. S. P. Nation & Coxhead, 2021). These figures are backed up by research that suggests that by the age of five, children are likely to know 3,000 word families, 4,000 by the age of 6, and so on (Biemiller, 2005). In 2001, Biemiller and Slonim reported mean normative vocabulary values of 5,200 root words by the age of 7–8 and 8,400 root words by 10–11, which also back up Nation's rule of thumb. Biemiller and Slonim used a recall interview test to measure the written receptive knowledge of word families in young first and second learners in English. In this format participants were asked to tell the interviewers what they knew about a word, with words scored as 'known' if participants were able to give an accurate definition. For very young learners, this can be problematic, and they are unlikely to get perfect definitions; for these, Biemiller and Slonim scored a word as known if the learner was able to provide some relevant information

about the word. Ensuring consistency for provided definitions and whether they qualify as 'knowing' a word is a problem with such interview tests.

Finally, we can look to studies of slightly older children to see if this rate of growth is sustained. Coxhead, Nation and Sim (2015) found that 17-year-olds nearing the end of secondary school had receptive vocabulary sizes of around 13,000 to 14,000 word families. After this time, vocabulary learning slows as speakers gain the vocabulary sizes necessary to meet their needs, and encounter fewer unknown words.

Table 3.1

Vocabulary Size Studies of Children in the Literature. NR = Not Reported.

Reference	**Age**	**Vocabulary size**	**Methodology**	**Words per year**	**Words per day**	**Notes**
Diller (1978)	7	50,000	NR	NR	NR	
Nagy & Herman (1987)	8–9	10,000	Recalibration of five older vocabulary studies.	2,000–4,000	8	
White, Graves & Slater (1990)	9–10	12–16,000	Multiple-choice test based on a random sample of words from the American Heritage Word Frequency Book.	3,300–5,200	NR	Advantaged populations knew more words.
Anglin (1993)	6–10	10,000–40,000 (or 6,000–20,000 root words)	Bespoke vocabulary test.	NR	NR	
Anderson & Nagy (1993)	16–18	40,000	Estimations drawn from schoolbooks.	2,000–3,000		
Biemiller & Slonim (2001)	6–9	5,200–8,400	Living Word Vocabulary lists and an interview format.	800	2.2	
Biemiller (2005)	8	5,000–6,000 word families	NR	NR	NR	
Coxhead, Nation & Sim (2015)	13–17	10–14,000 word families	20k version of the Vocabulary Size Test.	NR	NR	

3.1 Methodology

The original aim of this experiment was to gather longitudinal data on the receptive vocabulary sizes of a cohort of school-age children. This was intended to fill the research gaps in the literature regarding vocabulary development in primary school aged children in the UK; it would have also addressed the lack of research investigating growth rate. However, due to the COVID-19 pandemic and subsequent school closures, data collection could not proceed as planned and only one round of data was able to be collected in March 2020, just before the onset of the pandemic.

A large majority of research testing receptive vocabulary uses the Peabody Picture Vocabulary Test (PPVT) (Beglar & Nation, 2007; Dunn & Dunn, 2007), which is one of the most well-known vocabulary proficiency tests of American English. However, the PPVT does not provide an estimate of vocabulary size, but rather gives a standardised score that, out of context, does not provide deep information about a child's vocabulary. Scores from the PPVT cannot systematically represent a larger population of words (P. Nation & Anthony, 2016) and can only be compared to another learner's score; groups of scores can only be compared to other groups or norms for various groups.

However, what is more useful for vocabulary researchers are scores of vocabulary *size*, which can be gained from vocabulary tests such as the Vocabulary Size Test (Beglar & Nation, 2007). This test measures written receptive vocabulary knowledge of word families and has been used in several studies (Elgort, 2013; McLean et al., 2014). However, this vocabulary size test is not suitable for children, and is also not designed for native speakers. The Picture Vocabulary Size Test (PVST), by Nation and Anthony (2016), has been designed for younger, pre-literate first and second language learners of English. It is yet to undergo the rigorous validation that is necessary for the future of vocabulary testing research (Schmitt et al., 2020), but has been designed with a sound methodology, with a fair sampling rate and principled corpus selection (P. Nation & Anthony, 2016). The PVST uses

a computerised format and includes 96 test items drawn from twelve 500 word family lists (so a sampling rate of 8 items per 500 words) made from a specially created corpus of writing for children and adult spoken language, also checked against lists from a corpus of children's writing developed by Oxford University Press (Banerji et al., 2013). The test uses lists that follow a consistent word family definition, and so when this experiment refers to the number of "words" a child knows, it specifically means word families. This test was deemed suitable for the age range in question, is based on recent word family lists from a corpus of English relevant to children, has a decent sampling rate, and gives an estimation of vocabulary size that we can compare to the literature.

Figure 3.1 illustrates the setup of the Picture Vocabulary Size Test loading screen, which allows you to input the participant's name, gender, age, and also allows you to tag them for school or participant ID if desired. Figure 3.2 displays an example of the testing screen. Four pictures are displayed while the test plays an audio recording from a predefined script. For this screen, it would play "Along: Along there". The children were encouraged to point to the correct, matching picture. This continued for 96 items, and then the application provides a calculated estimated vocabulary size. There are two sets of wordlists (Appendix 1 and Appendix 2), and both were loaded into PVST and were used alternately in a split-half design. The children were tested individually in a small, quiet room. Testing time, on average, was around 15 minutes.

Figure 3.1

The PVST Loading Screen

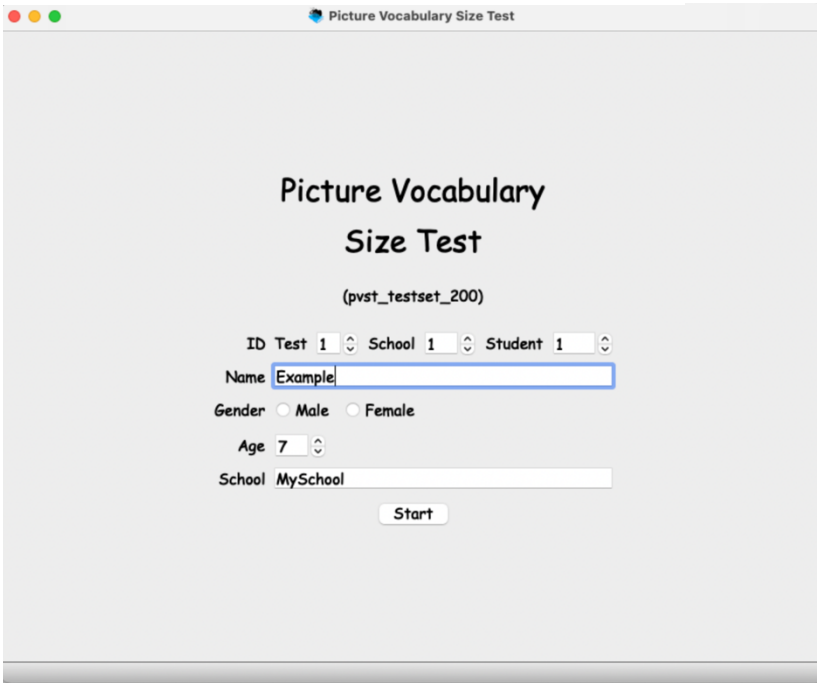
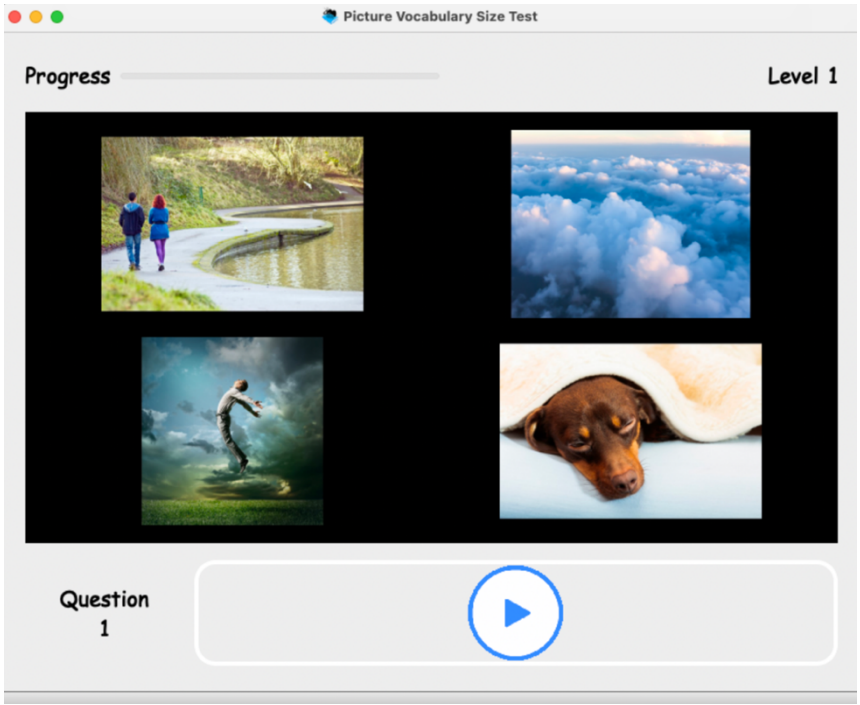


Figure 3.2

Example Testing Screen in the PVST



The school involved in this research was an English-medium school located in one of the least deprived areas in Swansea according to the 2019 Welsh Index of Multiple Deprivation (https://www.swansea.gov.uk/wimd2019 accessed 26 July 2024). The school has pupils from ages 3 to 11 and a total capacity of 216 children. The first round of data collection commenced on 26th Feb 2020 and finished on 13th March 2020. A total of three classes were tested: Reception (4–5 years old), Year 1 (5–6 years old), and Year 2 (6–7 years old). In this school, the Year 1 and Year 2 classes were taught together due to the very small number of children in Year 2, but for the purposes of investigation the groups were separated. There was a total of 40 participants.

Table 3.2

Participant Demographics for Children's Vocabulary Size Testing

	n	Female	Male
Reception (age 4—5)	14	7	7
Year 1 (age 5—6)	21	13	8
Year 2 (age 6—7)	5	3	2
Total	**40**	**23**	**17**

Scores and demographic information were collected via Excel and imported into SPSS for statistical analysis. Graphs were generated in SPSS. This study complied with the ethical considerations of Swansea University (ethical review number SU-Ethics-Student-170220/2417). Consent was given by the headteacher and all class teachers were given information sheets regarding the experiment. All children were sent home with opt-out consent forms to be filled in by their parents if they wished to opt-out of the research.

3.2 Results and Discussion

The results in this chapter showcase the vocabulary sizes of a small group of children aged 4 to 7 and suggest that children of this age know, on average, around 3,000 to 4,000 word families.

The scores exemplify the large amount of variation in vocabulary size present in this heterogenous cross-sectional sample. On average, across the 40 children tested in this sample, the children knew 3,503 words. The lowest vocabulary size measured was 2,000 and the highest was 4813 with a standard deviation of 571. There was an increasing trend in vocabulary size as children aged. The mean vocabulary size score for the 4-year-olds was 3,008, the mean vocabulary size score for the 5-year-olds was 3,441, the mean vocabulary size score for the 6-year-olds was 3,754, and the mean vocabulary size score for the 7-year-olds was 4,063.

Paul Nation and Avreil Coxhead state that young native speakers of English "begin school knowing several thousand words and increase their vocabulary size each year by close to 1,000 words" (2021, p. 43). In their examples of trialling the PVST in New Zealand, 6-year-olds knew around 3,500–4,000 word families and 8-year-olds knew 4,500–5,000. Table 3.3 reports the results of this study in comparison with to expected numbers of word families shown in research by Biemiller and Slonim (2001), Biemeller (2005), and Paul Nation, using his rule of thumb, i.e., vocabulary size of native speakers = (age – 2) * 1,000 (Nation, 2013; Nation & Anthony, 2016). The results in the study in this thesis suggest that children learn 300 to 400 word families per year rather than the roughly 500 words per year increase given by Nation and Coxhead (2021).

Table 3.3

Results Of Vocabulary Size Testing

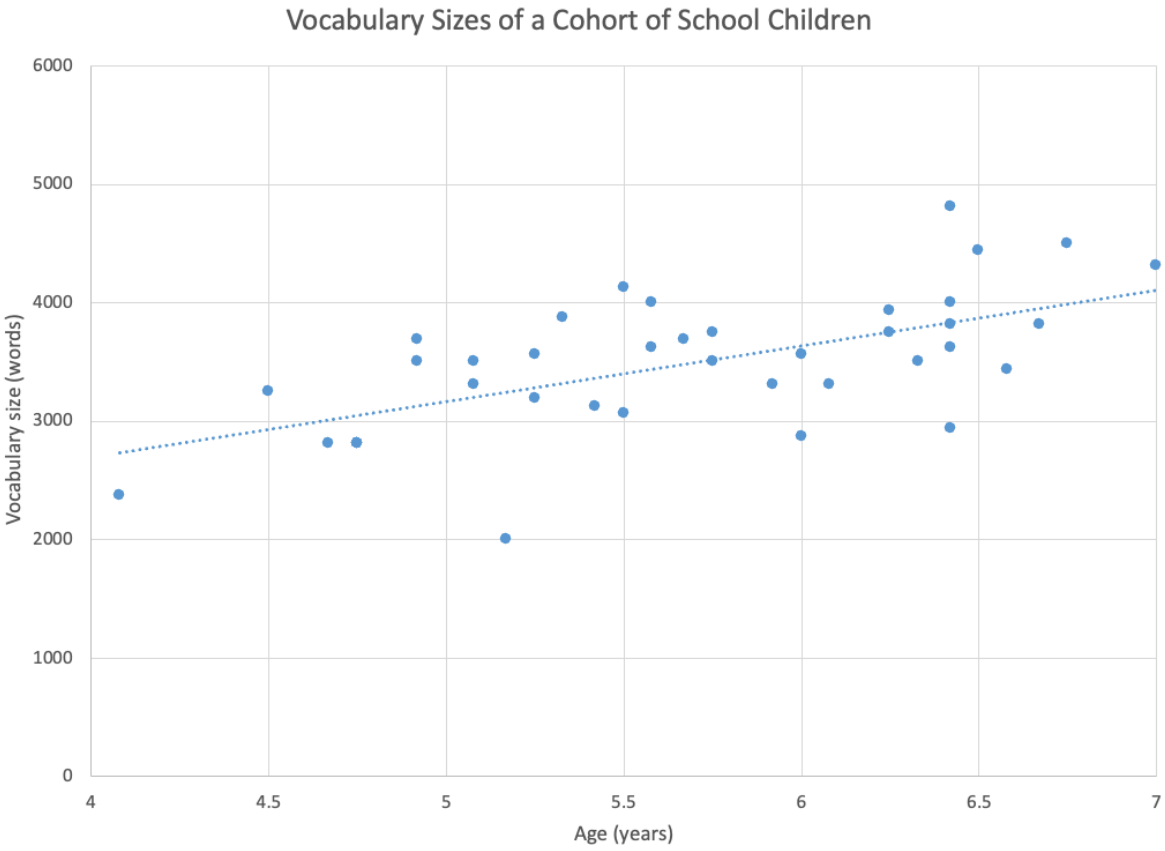
Age	**Expected number of word families**	***n***	**Mean vocabulary size score (word families)**	**Annual increase**	**Percent change**	**Number of words learned per day**
4 years old	2,000	8	3,008			
5 years old	3,000	15	3,441	+ 434	+ 14%	1.2
6 years old	4,000	15	3,754	+ 313	+ 9%	0.9
7 years old	5,000	2	4,063	+ 309	+ 8%	0.9

These numbers contradict the very large estimations of vocabulary size from previous research and suggest that children learn around one new word a day. The rate of learning, illustrated here by percent change, slows down as children age. The implications for this are that the vocabulary learning load may be lower than previous literature has suggested. Learning one new word a day seems like a much more reasonable endeavour than learning 5, 10, or 20 words per day needed to achieve the large vocabulary sizes measured in previous research.

Figure 3.3 shows a simple scatterplot of age versus score, illustrating the trend of increasing vocabulary size with age. A Pearson's correlation conducted in SPSS showed a modest correlation between age of the child and their score (r = .529, p = 0.01, 2-tailed). There is a lot of variation in these figures. For example, we can see that some older 4-year-olds demonstrated the same vocabulary sizes as some 6-year-olds. This is a result that reflects previous work which shows a wide range of sizes when you measure young children's receptive vocabulary, and that even the lowest scoring native speakers know several thousand words (I. S. P. Nation & Coxhead, 2021).

Figure 3.3

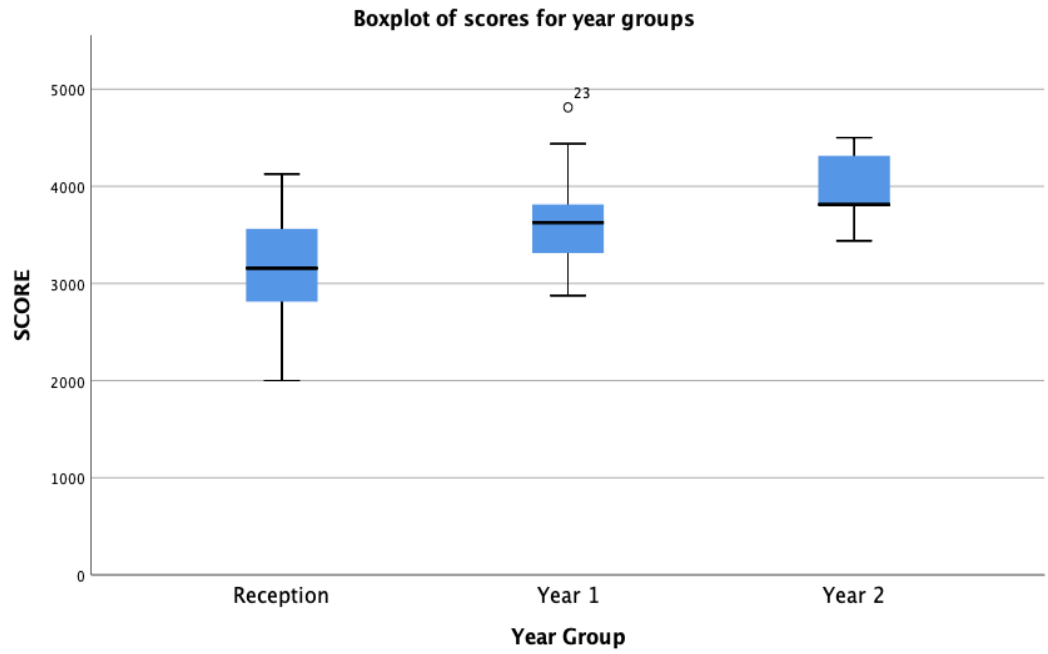
Scatterplot of Age and Vocabulary Size (Word Families)



We can also assume that vocabulary sizes will increase as we move up the year groups. A one-way ANOVA showed a significant difference between groups ($F(2,37) = 6.572$, $p = 0.04$). A Tukey post-hoc test showed that there was no difference in scores between Years 1 and 2, most likely due to the small sample size of Year 2. However, there was a significant difference between Reception and Years 1 and 2. If we illustrate scores for the three different year groups, we can see that there is less variation in Year 2, with Reception showing the most variation (Figure 3.4). However, this may, again, be due to the small sample size in Year 2. We can also see a general increase in scores as the year group increases.

Figure 3.4

Boxplot of Scores for Each Year Group



This chapter was originally intended to address the following research question: "How much vocabulary and which vocabulary is learned each year from three years' old up to the age of eleven?" However, due to the COVID-19 pandemic and its effect on the school systems, this section will mostly address the resultant limitations and discuss amendments to the research design which could be made to future work.

Data collection points were planned for 3- to 4-month intervals. However, the coronavirus pandemic and its subsequent school closures made in-person data collection impossible. Online testing was considered as an alternative, but was ultimately rejected due to difficulty in administering an in-person test remotely as well as safeguarding issues related to testing children remotely. What these results do show is that there is wide variation present in vocabulary sizes, as measured in this one sample with one test, and thus all classes have comparatively lower-

performing learners in the various age groups. It would have been important in future rounds to track these low-scoring students to see if they catch up to their peers. If they did not catch up, then it suggests that vocabulary interventions may be necessary. It is not clear if the vocabulary growth is as regular as numbers given by Nation and unfortunately, as the data collection was interrupted by COVID-19 pandemic, it was not possible to evaluate or model vocabulary growth.

Testing with the Picture Vocabulary Size Test was easy and children were engaged with the testing process. Limitations with the test instrument relate to test design and test-taking behaviour, which are common problems with testing of this nature. Children's behaviour was closely monitored and if guessing was suspected they were encouraged to carefully consider all options before making a choice. The PVST does include an 'I don't know' option, but this was not enabled in this study in lieu of careful monitoring of potential guesswork, and to avoid the temptation for children to repeatedly choose this option.

Referring back to the previously described discourse on the vocabulary gap, the sample size presented here was too small to draw any meaningful conclusions in relation to socioeconomic status. Only two children were eligible for free school meals (a common proxy for identifying disadvantaged pupils in the UK). Thus, there was no meaningful way to measure the impact of socioeconomic status on these results. Collecting a larger sample size and working with different schools would offer more opportunities to investigate the existence of any word gap.

Finally, it is important to highlight that this was only one test conducted with one sample testing one discrete area of vocabulary knowledge, that of receptive vocabulary. The results here do not give us any more information about what words children know and how they learn them. To investigate this, we have to turn to other methods. The following chapter will present the results of semi-structured interviews with teachers conducted after this vocabulary size testing took place.

4. The Impact of the COVID-19 Pandemic on Teaching and Learning in the UK

The COVID-19 pandemic severely affected countries all over the world and led to widespread restrictions on social movement and behaviour. In the United Kingdom, 'stay-at-home' lockdowns in response to the emerging crisis situation began on the 23 March 2020, and these included school closures in all four nations, leading to "the most disruptive period in children's education since at least the start of the Second World War" (Timmins, 2021, p. 4). Legal measures meant that all schools closed across the UK (with variations amongst the four nations), with the exception for provision for children of critical workers and vulnerable children (Sibieta & Cottell, 2020).

Thus, except for these particular cases, teaching in the UK became remote and home-based overnight (Welsh Government, 2021). Provision was largely organised by local authorities in Wales and Scotland, and by individual schools in collaboration with local authorities in England and Northern Ireland (Sibieta & Cottell, 2020). This situation persisted until early June 2020, when phased returns to school began. However, 'firebreak' and further lockdowns continued to disrupt learning. Table 4.1 summarises the restrictions and key dates affecting education throughout England, Wales, Scotland, and Northern Ireland in response to the pandemic.

Table 4.1

Key Coronavirus Lockdown Dates and Measures In The UK

Date	Description
23 March 2020	Prime Minister (PM) announces first lockdown in the UK, all schools closed
1 June 2020	Phased re-opening of schools in England
29 June 20202	Phased re-opening of schools in Wales
14 September 2020	Schools open to all pupils in Wales
31 October 2020	PM announces second lockdown in England
4 January 2021	Third lockdown announced; schools close across the UK
22 February 2021	Phased return of pupils began in Wales and Scotland, beginning with primary schools
8 March 2021	All primary schools open in England and Northern Ireland, secondary schools begin phased return
12 April 2021	All pupils return to schools in Wales
May–July 2021	Data collection for this chapter

This disruption is what motivated this chapter, which was not originally designed as a part of this PhD thesis, and instead was developed in response to the interruption of the study presented in Section 3, as the opportunity to draw out novel data of the impact of an unprecedented crisis on education in the UK could not be overlooked. Thus, this chapter aims to examine, through qualitative, interview-based methods, how teachers felt the pandemic and subsequent lockdowns impacted their teaching, in particular their literacy teaching, and how it impacted their students' vocabularies.

4.1 Literature Review

The school closures and learning-from-home measures in response to the coronavirus pandemic led to inevitable and difficult-to-measure impacts on children across the UK. Globally, school closures presented several challenges and consequences, including increasing inequalities in educational outcomes and widening of the learning gap between low- and high-income families (Van Lancker & Parolin, 2020). Soon after the first lockdowns began, copious amounts of scientific

research began to emerge on the impact of COVID-19 on people in all areas, from impacts on work, to health, to education. This extended to evaluating how the pandemic was affecting teaching and learning worldwide, as well as the impact of the pandemic on children's mental and physical wellbeing. As the pandemic unfolded, the effect of lockdown on literacy became a key topic discussed in the mainstream media as well as in academic publications, editorials, and commentaries (National Literacy Trust, 2020; Page, 2021; Woodcock, 2021). This literature review will discuss the pandemic's impact on two key, interlinked areas, with a particular focus on literacy at the primary school level: 1) the impact on teachers and teaching practices; 2) the subsequent effect of restrictions and new ways of teaching/learning on children's literacy development (contextualised within their overall school experience).

Global stay-at-home measures had a dramatic impact on education at all levels. Almost overnight, teaching in schools across the UK moved from traditional face-to-face learning in the classroom to an online interface which was at that point unheard-of in primary and secondary schools (Chapman et al., 2021; Lucas et al., 2020; Timmins, 2021; Waters-Davies et al., 2021). This emergency situation was unprecedented and, as such, complex policy and legislative decisions had to be made under compressed timescales without the usual impact assessment procedures or levels of collaboration (Welsh Government, 2020). In England, the Department for Education published regular guidance documents for schools and early year childcare providers. These have now been withdrawn but we can still access, for example, commentary on the initial guidance by SecEd: https://www.sec-ed.co.uk/best-practice/coronavirus-key-messages-dfe-guidance-schools-closure-key-workers-covid-19/ (accessed 24 January 2023). In Wales, education-related coronavirus legislation was issued from May 2020 with amendments to policies regarding attendance, the curriculum, and more, and archived versions of these documents can be accessed here: https://www.gov.wales/coronavirus-legislation-related-legislation#Educationandskills (accessed 24

January 2023). Scotland and Northern Ireland also issued their own legislations and had different lockdowns periods, but will not be discussed at length in this chapter.

In the UK, caregivers[2] were asked to mitigate impacts on their children by assisting them with home learning where they were able, and schools assisted with this by providing online resources (Sibieta & Cottell, 2020). However, there was no strict or dedicated guidance for this, and thus curriculum provision varied and the take-up of home learning was "uncertain" (Waters-Davies et al., 2021, p. 9). Teachers were tasked with adapting their ordinary lesson plans to new, online interfaces almost on their own, learning how to use new software while also ensuring their students could access and use the new software too. Meanwhile, some teachers remained in classrooms, teaching children of keyworkers and those who could not remain at home, based on government policy recommendations (Sibieta & Cottell, 2020). The transition to online learning presented significant challenges for both teachers and students, ranging from the practical limitations of adapting teaching to virtual learning environments (VLEs) to whether all students had the same opportunities to access online learning (for instance, did they have internet, a quiet place at home from which to work, the necessary devices, etc) (Chapman et al., 2021). Case studies, which are still emerging, describe how teachers across the globe adapted to the lockdown in different ways. In the United States, one report describes how teachers and students became proficient users of diverse learning environments as they transitioned to virtual learning (Chamberlain et al., 2020). The research from Chamberlain et al. (2020) highlights the ways in which they managed to make this

[2] Throughout this chapter, I will be using both the terms 'parents' or 'caregivers' to refer to biological parents, primary caregivers, or guardians of children.

transition more successful by focusing on maintaining a sense of community, “thoughtfully plan[ning] instructional strategies” (p. 244), strengthening parent–teacher communication, providing counselling services, and building community culture. They emphasise that by focusing on student wellbeing and providing natural, flexible opportunities to develop literacy virtually (through writing online journals, listening to teachers read stories, and being tasked with delivering daily morning announcements), literacy education was continually prioritised throughout the pandemic.

Governments were quick to commission research and inquiries into the pandemic’s impact on education and children’s services. Ofsted, the English school inspection authority, published a series of reports in 2020 which concluded that children of all ages and backgrounds showed evidence of loss of skills and learning due to coronavirus restrictions (Spielman, 2020). In Wales, the Welsh Government awarded grants to the eight Welsh universities in June 2020 to carry out collaborative research into the impact of the pandemic on the education system in Wales. They produced an empirical body of evidence, the main findings of which were that: (1) rapidly moving to blended/distance learning led to challenges, especially regarding access to the appropriate hardware/internet; (2) all learning was challenged by the situation, but some continued to thrive whereas others struggled especially those from disadvantaged backgrounds; (3) learners who thrived did so because of support they received from families, and learners who struggled did so due to missing the support of teachers and being unable to receive a high level of support from their family; (4) learners missed contact with teachers and peers, as well as specialist resources; (5) many learners struggled to progress their Welsh language skills. The collaborative group summarises their research by highlighting that the pandemic showed that many young learners are resilient and adapted quickly to changing situations. They also stress that “the lack of progress experienced by many vulnerable and disadvantaged learners was not a new phenomenon, but it was exacerbated

during this period and made more evident" (National Strategy for Educational Research and Enquiry, 2021).

Reports on the amount of home learning that took place during the coronavirus pandemic present significant heterogeneity based on the data collection process/survey, region, child, school, parental involvement, etc., and so reliable figures are difficult to calculate; however, the Education Policy Institute (an independent, evidence-based research institute that aims to promote high quality education outcomes for young people) in 2020 estimated that children were achieving, on average across the UK, 4.5 hours of home learning during the first lockdown period, which is a 25–30% reduction compared to pre-lockdown activities (Sibieta & Cottell, 2020). A similar figure was found in in that primary and secondary students were spending about 5 hours a day on average on home learning (Andrew et al., 2020). The Institute for Fiscal Studies highlighted that, almost a year after the start of the pandemic, total loss in face-to-face school time amounted to around half a normal school year for children across the UK and called for an impactful policy response to aid pupils in 'catching up' (Sibieta, 2021). The Education Policy Institute calculated that the lost weeks of schooling varied across the four nations depending on term times, exam leave, and the different re-opening times and procedures in England, Scotland, Wales and Northern Ireland; lost weeks of school thus ranged from about 5 weeks for those taking GCSE or A-Level equivalent exams in England, Wales and Northern Ireland, to 13 weeks for children in Scotland (Sibieta & Cottell, 2020). Other estimates put the total loss of face-to-face learning at 12 weeks for the majority of pupils across the UK in the initial lockdown period (Sharp et al., 2020; Waters-Davies et al., 2021).

This 'learning loss' has been a key theme discussed in publications related to learning during the pandemic. Although concerns have been raised about the impact of the pandemic on all areas of children's education, here we focus on literacy, e.g., the risk of losing reading and writing skills while not in school. As early as May 2020, researchers and organisations such as the National Literacy

Trust (2020) were highlighting the impact that COVID-19 disruption would have on the alleged literacy gap between advantaged and disadvantaged children (see Literacy Gap Chapter X). The impact of the pandemic on learning was a key area of discussion for news outlets worldwide, with dire headlines describing, for example, that the attainment gap in primary schools has widened since COVID ("Attainment Gap in Primary Schools Wider since Covid - BBC News," 2021) and claiming that 90,000 lack basic literacy because of COVID (Woodcock, 2021). However, it is important to highlight that the majority of literature published during the pandemic on this perceived 'learning loss' were only hypothetical or based on predicted losses, and the empirical research on the actual educational losses are scarce. What is now emerging is that learning losses are identifiable (Donnelly & Patrinos, 2022), but may not be as drastic as once thought (Gore et al., 2021). These impacts were not limited to the UK, of course. Language Magazine (Page, 2021), citing research from sources such as the United Nations, World Literacy Foundation, and Stanford University, highlighted that millions of children worldwide were negatively affected by the pandemic. For example, Merga et al. (2021) demonstrated the largely negative impact of the COVID-19 pandemic on writing instruction in Australian schools due to the disruption and necessary adjustments that had to be made to the curriculum; 72.6% of the 310 teachers surveyed felt that their writing instruction had been affected by the pandemic.

Learning losses have been found to be exacerbated by already present inequalities, such as the 'digital divide'. The 'digital divide' describes the gap between people who have full access to digital technologies (such as the internet and computers) and those who do not (C. Baker et al., 2020; Serafino, 2019). Concerns were raised throughout the coronavirus pandemic that this divide would be worsened due to increased reliance on an internet connection and personal computers or tablets to use in remote learning (Page, 2021). In the UK, disadvantaged children were also impacted by the pandemic to a greater degree than children who were not considered disadvantaged (Andrew

et al., 2020; Elliot Major et al., 2021), with research finding that pupil engagement was lower in schools with the highest levels of deprivation and that pupil access to IT and an adequate study space at home were significant challenges in the move to remote learning (Lucas et al., 2020).

However, not all the effects of the pandemic were necessarily negative. The National Literacy Trust released research in 2020 which showed that children and young people in the United Kingdom may have developed new and positive writing habits during lockdown (C. Clark et al., 2020). Their research was based on 4141 survey questionnaires sent to pupils aged 8 to 18. Increased enjoyment with writing was attributed to the increased amount of extra time which people had to write and think about writing. Many children used creative writing as an outlet for stress caused by the pandemic, with positive effects on their mental health. Some children expressed that they found greater benefit with writing digitally on laptops, but some said that the increased time spent on their computers for school affected their motivation to continue writing for pleasure on computers.

Technological innovations have had a great impact on literacy over the last decades. Using surveys conducted during the COVID-19 pandemic (between January and mid-March 2020, and between May and early June 2020) by the National Literacy Trust in the UK found that video games were a great source for literacy development for children, serving as a route into stories for young people and providing intrinsic motivation to read around wider topics related to video games (Picton, Clark & Judge, 2020).

Negative impacts of lockdowns extended beyond immediate learning needs. Mental health impacts due to lockdown restrictions and overall impact of coronavirus were raised as a key issue for adults and children, with the mental health charity *Mind* labelling the impact of the pandemic “a mental health emergency” (Mind, 2020). Student wellbeing was heavily affected as peer and teacher connections were suddenly removed and students had to adjust to new ways of working. However,

despite the initial panic about impacts of lockdowns on children's mental health, the findings emerging from the literature present mixed results, with some children struggling and some children thriving. Findings from the Children's Commissioner (2020) conducted in the very early days of the pandemic showed that children were beginning to experience stress regarding the pandemic, with 55% of children stating they had felt stressed (even if just a little bit) since their schools had closed; a higher proportion (61%) of girls felt stressed compared to boys (48%). Stress reduced, however, between their survey conducted in March 2020 and a follow-up survey in June 2020, which researchers attributed to a reduction in "extra little stresses" (p. 19) experienced at school, such as being told off by teachers. The key finding from these surveys was that, when asked what was contributing to their stress, responses in June revolved far more around coronavirus than responses in March did. Children expressed that they felt stressed about managing their schoolwork at home, about getting the correct grades, and about their family, friends and themselves catching COVID-19. Researchers also found that in March 2020, 75% of responses mentioned stresses to do with school, but by June 2020, this had reduced to 46%, with more stressors related to the pandemic being brought up. These results were echoed by a survey of 1046 Year 9 students conducted by the NIHR School for Public Health Research (Widnall et al., 2020), which found that, compared to pre-pandemic, there was an overall decrease in risk of anxiety and an increase in wellbeing, but that, when questioned, students had many worries related to COVID-19. The top three concerns were worrying about their friends/family catching the virus, worrying about the mental health of their friends and family, and worrying about missing school. The study also found that school connectedness (measuring a sense of school community and relationships with teachers) increased during the pandemic. A survey by Girlguiding (2020) found that positive aspects for the almost 7,000 girls they surveyed aged 4 to 18 included increased time spent with family, more time spent in nature, less academic pressure, and trying new hobbies.

The full impacts of the coronavirus pandemic and lockdowns are complex, multifaceted, and still emerging. The UK is now, at the time of writing, attempting to establish the 'new normal', and school has returned to a standard timetable and structure, with all pupils attending. There are many pandemic impacts that can be researched from both qualitative and quantitative perspectives. This research hopes to contribute to the ongoing emerging literature by qualitatively investigating the experiences during the pandemic of five schoolteachers in England and Wales, as well as on how they feel the lockdowns have impacted their students' literacy levels.

4.2 Aims and Methods

Thus, in order to contribute to the ongoing literature regarding the impacts of the pandemic on children in the UK, the overall aim of this study was to investigate how teaching and learning practices were affected during the pandemic and what impact these changes had on children's literacy development. This section introduces and discusses the methodological approach and research design used here in addressing this aim.

The specific research questions motivating this study were:

1. How did teaching and learning practices change throughout the COVID-19 pandemic?
2. To what extent do teachers feel their students have had their literacy development affected by the COVID-19-related government-mandated lockdowns in the UK?
3. What lasting impacts, if any, do teachers say the pandemic will have had on themselves and their students?

Qualitative semi-structured interviews are proposed as the best method to answer these questions. This section presents a justification for this decision alongside details of the key methods employed.

To answer these questions within the scope of this thesis, it was decided that several teachers would be recruited to complete semi-structured interviews to gather qualitative data. As the effects of coronavirus are complex and wide-reaching, interviews were chosen as an appropriate method of data collection. Interviews and focus groups have gained popularity in the social sciences in recent years, "partly as a consequence of a more general shift from quantitative towards qualitative methods" (Litosseliti, 2018, p. 156). In the field of applied linguistics, the use of interviews has also been increasing, particularly as a way to investigate participants' beliefs and experiences. Two reviews summarise the value of interviews and what can be learned from their application in the wider social science disciplines (Mann, 2011; Talmy, 2010). However, both these authors also highlight the fact that the interview is often under-theorised, inconsistent, and lacking both context and methodological detail. Talmy distinguishes between interviews as research instruments, which takes an intuitive and 'common sense' perspective on the interview, and interviews as social practice, where the interview is conceptualised and analysed as social action (Talmy, 2010, p. 129).

The advantages of interviews are that they are a useful and economical way to gain understanding of an individual's experience of a particular topic (Howitt & Cramer, 2007). They can be especially useful when researchers want to examine opinions, feelings, experiences, complex issues, and privileged information in more depth than can be gained using, for example, questionnaires (Denscombe, 2017, p. 203). Semi-structured interviews, in particular, allow for flexibility but with structure to ensure that research questions are addressed.

Alongside these benefits, challenges and issues have also been raised surrounding the legitimacy of data arising from interviews and focus groups. Bias is one challenge to avoid, and the interviewer must be mindful that they are not exerting an 'interviewer effect' on the participant, with also acknowledging that it is impossible for an interviewer to remain truly 'neutral' because

they bring with them their own view of the world. It is also important to acknowledge that when an interviewee agrees to an interview, they are agreeing to come to discuss a particular topic, and thus there is an element of 'manufactured' data, as argued by Silverman (2015). Ultimately, the interview is 'framed' by the theoretical underpinnings; the differences between positivism, naturalism, and constructionism approaches to interviews are discussed at length elsewhere (Silverman, 2015). Positivism suggests that interview data can provide us with facts about behaviour and attitudes, if the data is valid and reliable; naturalism posits than interviewees are subjects who actively construct their social worlds and thus interviews must be constructed to gain authentic insight into people's experience, e.g., through open-ended interviews; and constructionism proposes that interviewers and interviewees are simultaneously and actively engaged in constructing meaning together, and thus the topic becomes how meaning is being mutually constructed. Some of the above limitations can be avoided if, as is common in modern social science research, the researcher takes a constructionism approach by expecting that their participant will tailor their answers in response to the demands of the situation (i.e., respondents say what they believe or do, and interviews cannot offer insight into what they actually do) (Litosseliti, 2018). Constructionism also emphasises the fact that 'good' interview technique may matter less than some believe, as a successful interview depends on both interviewer and interviewee (Silverman, 2015). This approach corresponds to Talmy's definition of an interview as a social practice (Prior, 2018; Talmy, 2010).

The interpretation of meaning in interview-based research, and the theoretical views subscribed to, have implications for data analysis. The data analysis approach taken in this research was a thematic analysis, as the focus was on 'what' the participants produced (i.e., the stated experiences and opinions), rather than 'how' (Talmy, 2010). The analytic procedures are described in more detail below. Therefore, this study used interviews more as a research instrument than a social practice, and therefore adopted a more naturalistic approach. The way that meaning was

constructed and how knowledge was created were not analysed. However, throughout the interview design and implementation processes, the importance of bearing in mind constructivist views was maintained, with acknowledgment that data is collaboratively produced throughout the interview process.

3.1.1. Procedure

The participants of this study were recruited from May to June 2021 via a purposive sampling method. The research questions and aim of the study were to qualitatively evaluate the impact of COVID-19 and its subsequent lockdowns on teachers in the UK; thus, opinions had to be gathered from this population, i.e., teachers who were affected by lockdowns in the UK. In this manner, purposive sampling allows a researcher to seek out groups where the concepts being studied are likely to occur (Denzin & Lincoln, 2011). The final sample can be considered to represent the wider population of teachers across the UK who were affected by COVID-19 lockdowns.

A total of five schoolteachers were willing to participate in the study. Two were recruited from a local school in Swansea (Grange Primary School) via an email sent out by the head teacher in May 2021, inviting participants to express interest in this research study. Then, in June, an open call was posted on my personal social media accounts inviting any teacher to express interest. Seven teachers replied to this advertisement, and email addresses were gathered via personal message. Initial recruitment emails contained a description and explanation of the research as well as ethical statements and a consent form. Two participants were sourced from this recruitment. One was based in Kent and one in Surrey. The final participant was known to the department of Applied Linguistics in Swansea University and was approached personally; she was also based in Swansea.

Eligibility criteria were that they were currently employed as a teacher in the UK, they had worked throughout the COVID-19 lockdown periods, and that they fully consented to participating in the research. Ethical consent for this research was obtained from Swansea University (ethical

number). Table 4.2 presents the descriptive data of all five participants. The teachers all identified as female and were all from a White British background. Three taught in Wales and two taught in England. Their teaching experience ranged from 2 to 38 years. One school was a private institution and the rest were state schools; the private institution covered primary and secondary years, while for the remaining participants, one taught in a high school and three taught in primary schools.

Table 4.2

Semi-structured Interviews Participant Information

Participant ID	Years teaching experience	Year group(s) taught	Location
R1	3	Nursery, reception	Wales
R2	20	Year 1	Wales
R3	2	Year 6	England
R4	7	Years 7—11	Wales
R5	38	Years 5—8	England

Interviews were conducted via Microsoft Teams in May and July 2021 and ranged from 30 to 45 minutes. Interviews were recorded, with consent, and then were transcribed and imported into NVivo (Release 1.5.1). A total of 145 minutes of audio were recorded and the transcripts totalled 25,204 words. The interviews followed a semi-structured interview schedule that included key questions to examine how the participants felt the pandemic and its subsequent lockdowns had affected their pupils and themselves. The interview schedule is included in Table 4.3, listing the questions, probe (i.e., follow-up questions), and the intended rationale for each question. These questions were designed following the general interview guide approach, i.e., a semi-structured approach where the same general areas of information are collected from each participant, but with a degree of flexibility depending on how the interview develops (D. W. Turner, 2010). Specifically, this research followed the steps laid out by Bearman (2019, p.8):

1. Know your phenomenon of interest.
2. Aim for experiences more than opinions.

3. Start with a good warm-up question.
4. Brainstorm around the experiences you want to know about.
5. Use open-ended questions.
6. Consider the valence of your questions.
7. Leave space for interviewers to improvise; probes can help.
8. Start concrete and easy, finish with abstract and hard.
9. Final reflections offer opportunities for interviewee open comment.
10. Pilot, adjust the schedule and pilot again.

As Dörnyei points out, most interviews in applied linguistics are semi-structured as they offer a compromise between the two extremes of 'structured' and 'open' (Dörnyei, 2007). The interview guide in this research intended to strike this compromise, with a list of general questions intended to fulfil the research aim and optional follow-ups ('probes') to gather more information on research areas if deemed necessary. Semi-structured interviews are a particularly good way to elicit recounted events from participants, which Bearman (2019) feels are the key way to obtain 'thick, rich' experiential data. This type of data tells us much more about a participant's feelings and experiences and thus "a participant's recounting of events is the essential building block of qualitative interviewing" (Bearman, 2019, p.4). The questions in this interview schedule aimed to elicit this sort of data from participants by asking them to recount their experiences and describe their feelings.

Each question is presented with its rationale in the Table 4.3. In brief, four general questions started the interview, with the aim to understand how the participants were currently feeling and how they had felt throughout the pandemic, identify what had changed for them, and identify challenges they had faced. The interview started with an open question (framed as a statement, i.e., "I'd like to hear about your experiences over the last two years…"), as recommended by Richards

(2009). These questions were focused on gathering data to answer RQ1 (research question 1). Then, the focus of the interview shifts to asking literacy-related questions, such as how their literacy teaching practices had changed, how they felt their pupils' literacy had been affected, and if they felt that there would be any long-term repercussions on literacy development. These questions gathered data to answer RQ2. Question 8 was specifically geared towards answering RQ3. The questions were structured in such a way as to elicit experiences from the participants. Bearman explains that effective interview questions are those that generate rich descriptions and must contain meaningful prompts to evoke complex thoughts and descriptions, and that are relevant to the participants themselves, and which are clear and non-threatening (Bearman, 2019, p. 4). Open-ended questions are often the best way to generate these sorts of answers, as an open-ended question naturally elicits more information than a closed, yes/no questions, as open-ended questions minimise the number of predetermined responses, encouraging respondents to answer in their own words (Patton, 2014). The questions in the interview schedule are all phrased in a way to elicit open-ended answers, even if the first question seems to be closed (e.g., "4. Has the pandemic placed new or different demands on your role as a teacher?"), using follow-up probes if the participant were to answer the question in a closed manner (e.g., "How has the pandemic affected your role?"). This is called a closed–open pair and is common in qualitative interviews when an implicit expansion is assumed (Bearman, 2019). Every question included an optional probe or several probes to further encourage flexibility and "to guide the sort of improvisations necessary for the rich, thick description" (Bearman, 2019, p. 7). The final question was an opportunity for the interviewee to think about what has been said and add additional insights; the option for open comment is an important part of gathering final reflections, which can be a fruitful part of the interview process (Patton, 2014).

One limitation with such a semi-structured format such as the interview guide below, is that comparing participants and generalising findings can be difficult due to the lack of standardisation of procedures (Prior, 2018). For these interviews, every effort was made to ensure that all questions were asked, even if not in the same order. A limitation with this specific schedule is that, due to time and resource constraints, with the understanding that the more time that passed since the events of the pandemic, the less participants may remember about their experiences, a pilot questionnaire was not carried out, and thus the interview schedule questions, while refined, were not 'trialled' prior to being implemented in practice.

Table 4.3

Interview schedule including Question, Probe, and Rationale

Question	Probe/follow up	Rationale
Demographic information		
General questions		
Tell me about your experience of teaching over the last year during the COVID-19 crisis.	Best bit and worst bit	To gain a broad feel of their career stage, experiences, feelings, and mood
Can you sum up your feelings/experiences of the last year in three words?	Explain the words a bit more	To gather a series of words/phrases that sum up teachers' experiences over the last year
Have there been any changes (made by you or others) to how you have worked during this time? 4a. If yes: Can you describe these changes?	Who made these decisions? Who supported you to enact them? How have you managed the balance of teaching and other demands? What might be falling through the cracks as a result? (e.g., training, time, family, etc) Has the pandemic enabled you to do new things or do things differently?	To identify how people have had to change their priorities or usual teaching methods and balance competing demands, in general. To identify if new opportunities have arisen and what these are.
Has the pandemic placed new or different demands on your role as a teacher?	Have you had to step up to a new/different role? Tell us about this? How prepared did you feel for this? What support did you receive? Did you feel supported? How has the pandemic affected your role and how have you contributed? Where do teachers fit in?	To identify the challenges that teachers have faced and to explore whether they have felt supported throughout the crisis. To explore if teachers' roles have changed.
Literacy and vocabulary development specific questions		
How has your literacy-related teaching changed?	What have you prioritised and focused on in terms of literacy development? How have you or others decided where to focus your time and attention? What have you not done that you normally would, and what have you done differently/what is new?	To identify what teachers have deemed important specifically regarding literacy and teaching literacy over this year.

How do you believe children's literacy development has been impacted by the pandemic/closures/lockdown? 7a. If yes: What do you think are the main factors that have impacted children's literacy development and learning over this time (whether positive or negative)?	Can you give examples/expand on that? What has helped children succeed? And what are the barriers/what factors have hindered them or held them back? Are there any specific groups of children who have been disproportionately affected?	To identify to what extent teachers believe the pandemic/lockdowns/school closures have affected children's literacy and vocabulary development. To identify specific factors that have affected children's learning, whether positively or negatively. To identify factors that are out of the teachers' control, how important these things are, and to what extent they have affected children's learning.
Moving forward, how do you feel children will be impacted? What does the future look like, in the short-term and long-term?	Do you think the children will "catch up"? Will there be longstanding impacts, in your opinion?	To identify how teachers feel about the impact of the pandemic in the long-term.
Any other comments?		

4.3 Analytical Methodology

There are many different approaches to analysing the sorts of talk and texts that emerge when conducting social science research involving surveys, interviews, and focus groups, not only in applied linguistics, but in the broader fields of sociology, psychology, etc. The usual approach is to "reduce what the participants said to some sort of content categories" (Abell & Myers, 2008, p. 145) in a systematic fashion so that all sections are treated in the same manner, eliminating potential sources of distortion and attempting to establish a uniformity within the data. Although Abell and Myers were writing as supporters of qualitative discourse analysis (discussed below) rather than thematic coding, their point that this method is widespread in the social sciences still stands. The method used to examine the data in this study followed this sort of approach, in particular following a thematic analysis methodology.

The concept of thematic analysis covers a range of possible methodologies and the analysis employed in this research draws on work by Guest et al. (2014) and Braun & Clarke (2019). The former discuss applied thematic analysis, which requires a researcher to identify and describe implicit and explicit ideas within the data, based on patterns which emerge as themes throughout the data sets, then develop 'codes' to represent these themes and which are used for later analysis (Guest et al., 2014, p. 9). It is an exploratory approach that generates codes derived from the data which are not pre-determined, and moves beyond merely counting the words or phrases. Braun and Clarke (2019) introduce reflexive thematic analysis, which places emphasis on the researcher's "reflexive and thoughtful engagement with their data" (p. 14). Thematic analysis is suitable for analysing large data sets and allows researchers to identify and examine nuances in the data that mimic real life discussions (Braun & Clarke, 2013; Denscombe, 2017), thus allowing a researcher to find patterns to understand the phenomenon under investigation. Above all, Braun and Clarke

emphasise the importance of approaching qualitative research in a creative and thoughtful way that considers subjectivity as a valuable resource:

> We intended our approach to TA to reflect our view of qualitative research as creative, reflexive and subjective, with researcher subjectivity understood as a resource (see Gough & Madill, 2012), rather than a potential threat to knowledge production, as it arguably is conceptualised in Boyatzis' and some other approaches to TA. For us, qualitative research is about meaning and meaning-making, and viewing these as always context-bound, positioned and situated, and qualitative data analysis is about telling 'stories', about interpreting, and creating, not discovering and finding the 'truth' that is either 'out there' and findable from, or buried deep within, the data. For us, the final analysis is the product of deep and prolonged data immersion, thoughtfulness and reflection, something that is active and generative. (Braun & Clarke, 2019, p. 7)

The ideas of thematic analysis, discourse analysis, and grounded theory are often conflated; grounded theory is defined as a set of techniques which identify categories and concepts within text which are then linked to formal theoretical models (Corbin & Strauss, 2008; Guest et al., 2014). Thematic analysis is linked to grounded theory as is also has an emphasis on supporting claims with data (Guest et al., 2014). The difference is that, where grounded theory builds theoretical models from the data, applied thematic analysis does not require a model to be an output of the process. What remains important is that our interpretations are supported by real data.

The largest limitation with this method is reliability, as the method is subject to interpretation, bias, and subjectivism; however, it is still a preferred method among social science researchers using qualitative interview data as it is systematic, yet flexible and inductive (Guest et al., 2014). And, as Braun and Clarke highlight, subjectivity is not necessarily a detriment to qualitative research. This emphasis on context is echoed in the theory on qualitative discourse

analysis, which is a key methodology in applied linguistics when studying any sort of discourse, including research interviews, focus groups, etc. Researchers in qualitative discourse analysis highlight the importance of analysing research interviews in their intertextual, interdisciplinary, socio-political and immediate contexts, taking into account sociological variables, the history of an institution, and the institutional frames of a conversation to reveal the various strategies and structures within different levels of text and talk (Wodak, 2008). While these methodologies and sub-disciplines (e.g., conversation analysis) have brought an undeniable level of rigour to the field of discourse analysis and applied linguistics as a whole, the level of analysis involved in these fields that would look at co-text, concordances, intertextual links, context and more in research interviews (Abell & Myers, 2008) is beyond the scope of this chapter, and is not of interest to this researcher (with focus rather placed on the actual things my interviewees expressed to me, rather than how they said it). However, I do acknowledge the importance of considering all the utterances in these interviews as rooted in several contexts: the time and place (i.e., post-COVID, online), the understanding of the purpose of the interview, and the relationship between the interviewer and interviewee.

Once interviews had been conducted, the data analysis followed procedures outlined by Denscombe (2017) and Braun and Clarke (2013). In summary, this process included: (1) Produce verbatim transcripts; (2) Identify possible themes; (3) Then, the transcripts were carefully read and the data were checked (e.g., correcting typos, ensuring the transcripts were accurate by cross-referencing with the audio files); (4) Import the transcripts into the appropriate software. In this research, the transcripts were imported into NVivo (v.1.6.1). (5) Read and re-read the transcripts as many times as necessary in order to identify common themes using NVivo's 'codes' function. The resulting codes are presented in the following results section.

Bringing this all together, as an example for how the thematic analysis occurred, consider the following decontextualised statement:

> "So, using Zoom as our main platform for delivering this, and we did start off using Microsoft Teams for some of the kids, but it was more complex for them." (R5)

At first glance, this statement seems to be discussing themes of technology and someone's view of how some children coped with a situation. But these decontextualised ideas are not enough — we must consider the context of the situation and the interview prompt that proceeded this statement. In this manner we can format the extract as below, presenting the preceding statement with line numbers which align with the transcripts. The transcripts are not included in this document due to their length (approximately 25,000 words) but can be provided upon request. Short excerpts will be referred to in brackets, followed by the line number, but longer extracts will be formatted as below:

Extract 1

Speaker	Line #	
Interviewer	27	So just tell me about your experience of teaching during COVID and the
	28	pandemic and everything that's been going on
R5	34	So, using
	35	Zoom as our main platform for delivering this, and we did start off using Microsoft Teams for
	36	some of the kids, but it was more complex for them

This places the response in more context, namely that the interviewee is summarising a change or adaptation made in response to the COVID-19 pandemic. For this extract, once we understand the context of the situation (Abell & Myers, 2008, p. 154), i.e., upon knowing the interviewee is a teacher, we can surmise that the "delivering this" in the above statement refers to delivering a lesson, i.e., teaching. We can infer that the lessons had moved online when we understand what sorts of platforms Microsoft Teams and Zoom are (i.e., online communication

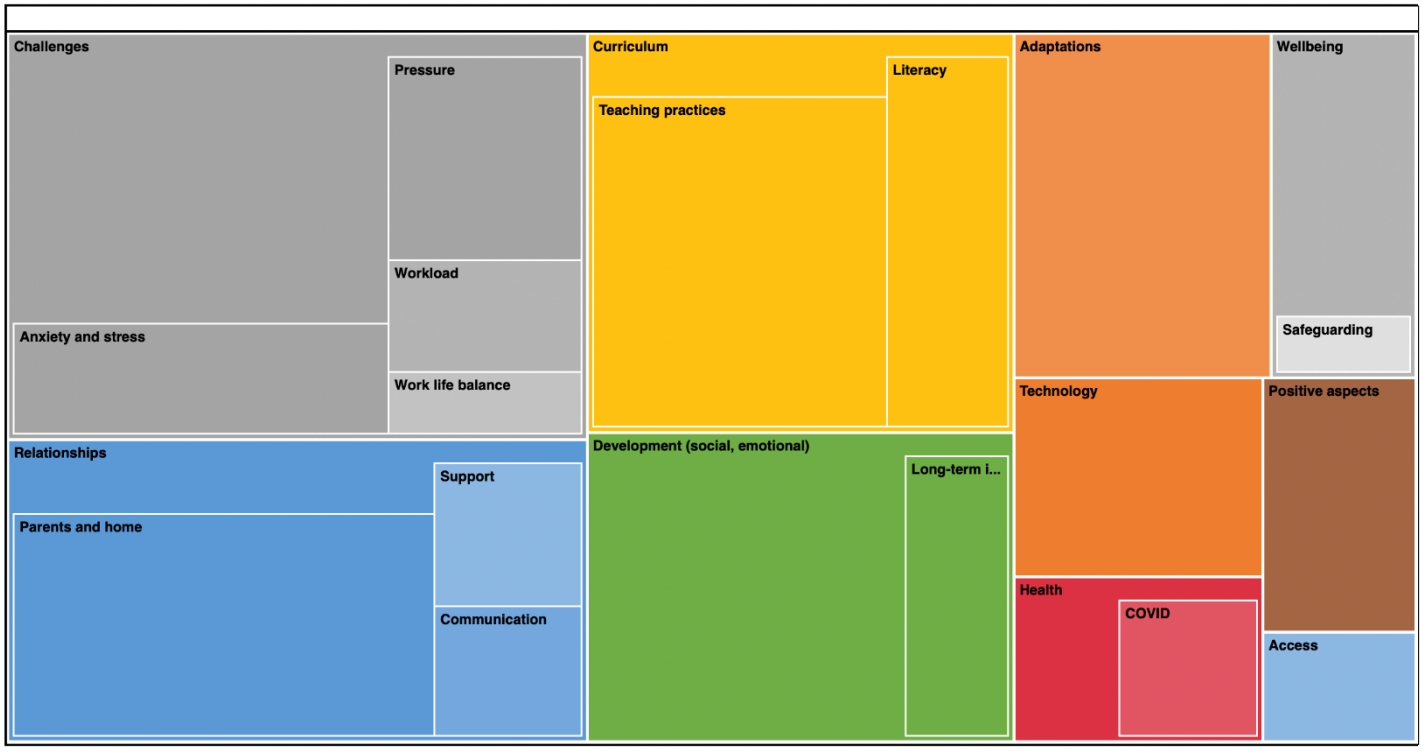
platforms) and we are made aware of the changes that the COVID-19 pandemic and subsequent lockdowns led to in terms of remote teaching and learning. Suddenly, when this seemingly simple statement is more carefully considered within its larger context, and compared against the rest of the interview and the other interviews, we can decide that the actual key themes for this statement are: (1) adaptation/changes (to teaching strategies in response to COVID); (2) technology; (3) challenges (for teachers and students). This example demonstrates how codes often overlap, or it is difficult to decide what theme an utterance was reflecting. Thus, one statement may be coded with two or even three themes. This allows us to identify patterns, e.g., changes made during COVID-19 were often challenging. While seemingly an obvious idea, something that could have been hypothesised prior to this research, the fact that these connected ideas emerged through an analysis of these data without preconceived notions is extremely valuable. It was in this manner that all five interviews were subsequently analysed and coded based on the methodologies of thematic analysis and considering the importance of context discussed in qualitative discourse analysis research.

4.4 Results, Analysis and Discussion

All respondents spoke about the impact the pandemic had had on them, their colleagues, and their pupils. Following the procedures outlined in the Aims and Methods, a total of ten codes were identified as top-level, with some sub-codes grouped together under these umbrella headings for better organisation (e.g., first pressure, workload etc. were coded separately, then grouped under the subheading 'challenges'). Thus, the final codes were: (1) access; (2) adaptations; (3) challenges (sub-codes: pressure, anxiety and stress, workload, work–life balance); (4) curriculum (sub-codes: teaching practices, literacy); (5) development, i.e., social, emotional, and educational (sub-code: long-term impacts); (6) health (sub-code: COVID); (7) relationships (sub-codes: parents and home, support, communication); (8) positive aspects; (9) technology; (10) wellbeing (sub-code: safeguarding). Figure 4.1 presents a treemap chart visualising the number of coding references for each of these parent codes and their child codes.

Figure 4.1

Treemap Generated By Nvivo where Size Indicates Number of Items Coded for that Label



This treemap shows us the distribution of the codes in this dataset. The squares are organised by size, with bigger squares indicating that more items were coded for that label. We can see that *challenges* was the biggest code, followed by *relationships*, and more items were coded for *parents and home*, indicating that parents were a frequent topic of discussion from these participants.

After the coding step was completed, a final inductive step was carried out to evaluate the overall narrative of the data. Looking at the datasets all together, three overarching themes were discovered: (1) how the pandemic changed teaching and learning practices for the teachers (Section 4.5: Impact on Teaching); (2) the identified importance of external support for children (Section 4.6: The Role of External Support); and (3) the pandemic's impact on wellbeing for both students and teachers (Section 4.7: Impact on Wellbeing). The data in this chapter are presented under these overarching themes, with all themes drawing evidence from the identified codes to form the narrative. It is not possible to say that each code is only linked to one theme, as many of the codes cut across themes (for example, the theme Impact on Teaching draws on the codes challenges, curriculum, adaptations, etc); the themes are thus a method to narrativize what was found in the dataset. Table 4.4 presents a summary of the codes from the teacher interviews, alongside contextualising examples from the data.

In general, the interviews showed that the coronavirus pandemic and its subsequent lockdown restrictions affected children's learning and wellbeing to different extents. Some children thrived, while others struggled. These interviews showed that the teachers felt that the children were not only affected academically, but were also sensitive to the changes going on in their lives and were affected emotionally and mentally, even in the younger years. Staff were also affected, with many different demands placed upon them during the pandemic. The amount of external

support that children could access was noted as a key aspect that affected children's learning outcomes. However, as this research did not investigate the direct links between external support and learning outcomes, we do not know if there is a causal relationship. What we do know is that the ideas that emerged from these interviews are reflected in academic and governmental research which will be presented in this section alongside the results from the interviews.

Table 4.4

Summary of Codes from Teacher Interviews

Code	Summary	Evidence
Access	Not all children had access to the required technology to work from home; some schools offered laptops or tablets to students who needed them, but not all could provide this. Some schools delivered printed handout packs to those who needed them.	"Just certain children, those who were really struggling and who didn't have any sort of access to device or really weren't able to use it that often." (R3, line 1042)
Adaptations	Teachers had to quickly adapt to changing circumstances using tools which most had not used before. Some struggled with adapting to the new interfaces and some of the new technologies led to anxieties in respondents.	"So obviously that that was when we were all suddenly thrown into this. We don't know what we're doing. We had to start learning how to use [Microsoft] Teams and things like that and teach the children how to use it with lots of uncertainty around that." (R3, line 980)
Challenges (including pressure, anxiety, workload, and work–life balance)	The most discussed challenges were related to adjusting to continuously changing circumstances; dealing with the transition periods between lockdowns; and increased pressure and workload on teachers.	"Okay, so I've got a class which has got quite a lot of needs so there were quite a lot of children who struggled over the initial lockdown period, they missed a lot of year five. And so coming back in September, it was quite difficult for a lot of them. In terms of how much learning they'd lost." (R3, line 909)
Curriculum (including changing teaching practices and literacy)	There were different approaches to timetables and delivering pre-recorded versus live lessons. Some schools adapted their curricula depending on what students could handle, while others endeavoured to keep to the same curriculum. All respondents felt there would be some long-term learning loss, but they were hopeful the majority of their students would 'catch up'.	"So we would set assignments using [Microsoft] Teams, so we would do a maths and then English assignment and we'd do set times tables and spelling quizzes." (R3, line 979)

Development (sub-theme: long-term impacts)	Social, emotional, and educational regression was observed, especially in the younger years. Anxiety upon the transition back to the school environment was a common theme.	"But from an educational perspective, they progress, I feel, so much more when they're actually in that classroom experience, in that classroom education environment. It's that there's a mind switch, to 'okay, this is what I've got to do now'. Whereas at home, it's like 'I haven't got to do this though.'" (R2, line 404)
Health (sub-theme: COVID)	Respondents discussed their increased focus on hygiene upon return to school and mentioned personal brushes with COVID-19. COVID 'bubbles' and isolation were also mentioned throughout the interviews as new challenges that were dealt with.	"We used to like chuck the PE mats on the floor and then chuck them in the corner and chuck them out again. And they were never wiped down or, you know, so I think people's awareness of such as some really basic things like that is much more, you know. We're much keener, really, on everything." (R5, line 2143)
Relationships (sub-themes: support, communication, parents and home)	Teachers remarked on the importance of staying connected with their colleagues in order to support one another. Parent–teacher relationships evolved, and parental input was highlighted as a key aspect that contributed to how students progressed during lockdown periods.	"Some parents haven't done anything at all. So then obviously we, we come back, then we've got the children who've come on massively because the parents have been doing everything and more with them. And then we've got children who have literally done nothing. And then it's like starting from the start then." (R1, line 71)
Positive aspects	Respondents noted that staff and students demonstrated remarkable flexibility and willingness to overcome challenges and try new things. An increased sense of resilience was observed. Some staff enjoyed working from home. Respondents expressed pleasure at returning to the classroom post-pandemic.	"But coming back then, it's just been amazing. It just, it just means so much. They've had that time not being in school. And then seeing them in school and just seeing how much more they get from the whole experience of being there." (R2, line 474)
Technology	Technological innovations and learning to use new technology were at the forefront of the pandemic. Microsoft Teams, Zoom, and Seesaw were the primary applications identified in this research.	"We used Seesaw as an engagement tool. So I posted some activities online and the children would be able to do them and then post back and things like that." (R2, line 450)
Wellbeing (sub-theme: safeguarding)	Teacher wellbeing was significantly affected by an increased workload, blurring of the lines between work and home, and decreased work–life balance. Student wellbeing was affected by the stress and upheaval of the year and by	"You really need is time to be able to talk to those parents, build those relationships up, have sessions with the children about their wellbeing and have that as the focus rather than the focus being on the academics at this stage

	the transition periods between remote teaching and in-classroom learning.	because they've had such a, it's been such a difficult time for them." (R3, line 1226)

4.5 Impact on Teaching

This section will explore the complex and varied impacts mentioned by respondents that the pandemic had on teachers and students specifically in regard to their teaching and learning, covering aspects such as the adaptation to remote teaching, the implementation of new technologies, how schools approached the need to adapt their curriculum and timetables, and the perceived impact on the students' literacy. All respondents described how the lockdown had affected their teaching as they all moved to a fully remote and/or blended teaching experience (fluctuating over the duration of the pandemic depending on local lockdown rules). Approaches to the curriculum varied, with some schools adopting a more flexible routine and others trying to keep things as much the same as possible. As teaching moved online, different schools trialled different strategies, trialled different technologies, and implemented varying policies; these are summarised in Table 4.5. The particular technologies are described further in this chapter where relevant. The results and discussion in this section answer both research questions.

Table 4.5

Approaches to Online Learning

Participant	Approach to online teaching
R1	Tried to keep things "as normal as possible" Microsoft Teams meeting once a week No big writing topic; focus on skills-based work Seesaw, Microsoft Teams, Reading Eggs Pre-recorded lessons
R2	Seesaw as the main engagement tool Microsoft Teams meetings started off every day, now once a week More prescriptive approach Pre-recorded lessons
R3	Microsoft Teams with class notebooks where activities were placed Teachers were given specialist subjects to teach across three classes Children were called by phone once a week Pre-recorded lessons at first, then later trialled live lessons
R4	Timetable did not change and was streamed live all day Emailing, email groups on Hwb (a digital learning platform used by schools in Wales) Later started Instagram and Twitter pages to keep everyone connected/informed Mini whiteboards, Quizlet, Kahoot! Live lessons
R5	Zoom as their main platform, with a camera-on policy, and also used Microsoft Teams Adapted the timetable so that there were fewer lessons, but lessons were longer (an hour for each session) All afternoon sessions were creative or sports-based Continued with some extra-curricular activities online e.g., linking up children playing music to produce orchestral pieces Live lessons

4.6.1. Adapting to Remote Teaching

This sudden shift to online teaching presented a dramatic challenge for schools across the UK as they scrambled to adapt to this emergency. Whereas previously institutions may have had crisis management protocols in place in terms of, for example, school violence or sudden natural disasters, i.e., sudden impact events with a defined beginning and end, the COVID-19 pandemic

presented a gradual, drawn-out crisis, also referred to as a “creeping crisis” or “compound crisis” (Hulme et al., 2021). This event necessitated a high level of collaboration, preparedness, leadership, flexibility and compassion and research has highlighted the wide range of adaptive leadership strategies employed by headteachers across the UK (Hulme et al., 2021). Teachers, as leaders in their own right in their classrooms (Lieberman & Miller, 2005), were key in this adaptation. The move to online teaching was primarily supported through several different tools, applications, and websites that schools utilised in their teaching. This section will summarise ‘how’ the respondents teaching changed during the pandemic as they adapted to remote learning (and the following section will cover the ‘what’, i.e., their delivery of the curriculum and educational content).

Online teaching and learning is not a new concept, particularly in higher education (A. Sun & Chen, 2016), and many terms are used to refer to this style of teaching to differentiate it from traditional face-to-face teaching and learning: the literature refers to numerous combinations of the words *digital*, *online*, *e-*, or *virtual* with the terms *remote*, *distance*, *home* and the nouns *education*, *learning*, *teaching*, *instruction* or *schooling* (Coleman, 2021), which makes a review of the literature on online education challenging due to the variable uses of these terms. The term *remote learning* was not widespread in the literature until the COVID-19 pandemic; *emergency* then began to be added into the mix (Barbour et al., 2020). This new term ‘emergency remote teaching’ has been defined as “a temporary shift of instructional delivery to an alternate delivery mode due to crisis circumstances. It involves the use of fully remote teaching solutions for instruction or education that would otherwise be delivered face-to-face or as blended or hybrid courses and that will return to that format once the crisis or emergency has abated.” (Hodges et al., 2020, p. 7). This reflects the position that online teaching held in the education sector in the UK during the COVID-19 era.

Following a review of 134 empirical studies published prior to the pandemic, Carrillo and Flores (2020) concluded that an effective online educational experience is derived from several

interlinking factors, including teaching presence (e.g., pedagogical approach, learning design), social presence (e.g., participation, interactivity), and cognitive presence (e.g., contextual factors). Their literature review highlighted that the "active and thoughtful participation of all the members involved in the online learning environment was core to the achievement of effective practices" (Carrillo & Flores, 2020, p. 477), and that supportive interactions are needed to mediate the learning process across the internet. The important contextual factor relevant to this research was the emergency nature of the shift to online teaching, and thus the lack of preparation and knowledge of how to transition to remote learning. Teachers were not able to efficiently capitalise on the extensive research carried out on online learning or receive adequate training due to the sudden and emergent nature of the pandemic, as has been highlighted in prior research (Noor et al., 2020).

The shift to online learning began by identifying the appropriate remote tools. Microsoft Teams was referenced by all participants as a main application used to adapt their teaching to the online sphere. The video-conferencing platform Zoom was also mentioned by one participant (R5). Microsoft Teams and Zoom are both platforms that facilitate one-on-one or group video calls, with Microsoft Teams in particular a successor to the video-calling application Skype, but with the added benefit of further features that were capitalised on during the pandemic to facilitate online teaching and learning, such as being able to being able to set and mark assignments in the platform. While both applications existed before the pandemic, they had a noticeable increase in users during 2020: from March to June 2020, Microsoft Teams had a growth of 894%, and as of 2023 over 183,000 educational institutions are using Microsoft Teams; Zoom had a similar boom in use during the pandemic, and at its height over 90,000 schools were recorded to be using Zoom (Curry, 2023). Many of the teachers surveyed in this research discussed their experiences of using these applications for the first time, highlighting that both staff and students had to get to grips with a myriad of new software (Extract 2).

Extract 2

Speaker	Line #	
R3	980	So obviously that that was when we were all suddenly thrown into this. We
	981	don't know what we're doing. We had to start learning how to use Teams and things like
	982	that and teach the children how to use it with lots of uncertainty around
	[…]	that. […]
	1004	We were suddenly having to learn how to do all these things and there was still the expectation that
	1005	we were marking everything, that we were then phoning children.

As the pandemic unfolded, managing these new technologies became easier ("But second lockdown, it was a little bit easier in terms of knowing what to do, because we set up Teams […]" (R4; line 1320)). These applications were the main way in which schools managed their curricula and timetables, but were used to varying degrees. Some schools endeavoured to keep their teaching "as normal as possible" (mentioned by R1, R2, and R4) and tried to emulate a normal school day as much as they could on these video conferencing apps. R4 explained that their timetable was kept essentially the same and was streamed live all day on Microsoft Teams (Extract 3).

Extract 3

Speaker	Line #	
R4	1321	The school decided we needed to follow the
	1322	full timetable on Teams so we were streaming then registration at 8:40 all the way through
	1323	then to 3:10, live to everybody. Students weren't allowed to have cameras and mics on,
	1324	so they were just communicating via chat and we were just basically talking through the
	1325	work, trying to make it as engaging as possible. And then, we were tracking them, with the
	1326	assignments you can see on Teams who did what and how long they were in your
	1327	lesson for. But to be honest, it was stressful because children weren't engaging, they were
	1328	just sitting on, not joining in the lesson.

This quote illuminates several salient points which will be further explored in this section: the challenges of engaging students online, the quandary over whether to deliver live or pre-recorded content, and the decisions that had to be made over certain policies (e.g., camera-on or camera-off, mic-on or mic-off). Adapting to online teaching and learning came with many difficulties for both staff and students and all the participants touched on these challenges and difficulties in managing remote learning compared to in-person delivery. For younger children, difficulties revolved around learning to use the software and managing distractions; for older children, there were more problems with students turning off their cameras, muting themselves, or not completing work. For example, R5 described problems with their 'camera-on' policy: "We had a camera-on policy. Some schools went for camera-off policy. We went for a camera-on, we wanted to see the kids. But the older children sometimes were very reluctant to do that." (R5; line 1975). Camera-on versus camera-off policies can affect both engagement and connectedness during remote teaching and learning, with educators having to manage the fine line between ensuring their students are 'present' and engaged in their lessons, and not pressuring or making students feel self-conscious if their cameras are always turned on (Martin & Bolliger, 2018; Schwenck & Pryor, 2021). Comparing Extract 3 with this quote from R5, we can see that problems were encountered no matter whether the schools implemented a 'camera-on' or 'camera-off' policy.

Rather than delivering live lessons, R1 and R2 described how they delivered pre-recorded lessons, utilising Microsoft Teams and Seesaw as their primary engagement tools. Thus, although they said that they endeavoured to make their teaching as normal as possible, the difference between those who streamed lessons live was the pre-recorded and thus asynchronous nature of the teaching. Some schools trialled mixed approaches; for example, R3 described how during the first lockdown they did no live lessons but set assignments, spelling quizzes and worksheets using class "notebooks" in Microsoft Teams. Lessons were pre-recorded, which she described as time-

consuming, as often you would have to do five or six takes due to mistakes and, thus, a ten-minute lesson might take over half an hour to pre-record. In subsequent lockdowns, they trialled live lessons over Teams followed by tasks, which worked better. R2 also described how pre-recording lessons was extremely time consuming and described how she kept deleting and restarting video lessons as she felt a lot of pressure and anxiety to get it 'right' (Extract 4).

Extract 4

Speaker	Line #	
R2	647	we had to, not just me, do a lesson starter and that would
	648	be a video for one or two of the lessons that you're posting for that day. And that
	649	I thought, I was anxious about that because when you're in the classroom with the children, you can
	650	talk, they know your way, you know their way and they know, you know, what they respond
	651	to. But because you've got your online lesson and parents can hear you, that caused some
	652	anxiety within me and I don't know how many times I started a lesson, deleted it, started a
	653	lesson, deleted it, it used to take me all day. So that was quite of an anxiety with me, really.
Interviewer	655	Well, it's a completely different interface, isn't it?
R2	657	Yeah. Yeah, it really is. And sometimes you said, my gosh, have I said the right
	658	thing, have I used the right words, have I been engaging with the children? But in the end I
	659	just had to post something. Because that was the expectation. But it used
	660	to take me all day to think about one lesson starter.

These pressures and anxieties ("expectation", as said in Extract 4, line 659) to produce content they had never had to produce before (line 657) contributed to the increased workload felt by several participants and which is further discussed in Section 4.2. The participants also discussed how it was difficult to manage parents' expectations, as the parents were also getting to grips with the new software and systems (e.g., *"I had some messages from parents saying I'm really struggling to keep up with the tasks that you're posting online" (R2; line 633)*). Parental involvement

throughout the pandemic was a key theme in these interviews and is discussed in more depth in Section 4.2.

In contrast to those who tried to keep their timetable similar to an in-person timetable, R5 described how their school adapted the timetable quite a lot, so they had fewer but longer lessons and afternoon sessions were dedicated to creativity or sport, and they used Zoom as their main platform and delivered live lessons. They offered a level of flexibility to their children based on what they could manage (Extract 5).

Extract 5

Speaker	Line #	
R5	1614	Obviously, we adapted the timetable significantly. So they didn't have exactly the
	1615	same number of lessons. They had fewer lessons, but they were for a longer period of time.
	1616	And so roughly about an hour for each session. And I'm talking only about the children from
	1617	year 5 to 8 at the moment. And then we, we made all the afternoon sessions were creative
	1618	and sport based. So some kids actually couldn't manage to do a full day of online learning at
	1619	home and they would dip out after lunch time. We're very flexible and we adapted to the
	1620	needs of children on an individual basis because some of them couldn't manage things like
	1621	French when they were working at home.

These varying approaches to delivering remote teaching throughout the pandemic are reflected in the literature. In their COVID-19 schools briefing series, Ofsted described how the schools they visited (a total of 121 pilot visits) were “generally using online methods of delivery – recorded online lessons, individual study modules or often a combination of the two. Occasionally, leaders said they were using live online lessons.” (Ofsted, 2020a, p. 4). A report one month later described how some schools had reservations about running live lessons, due to potential safeguarding issues (e.g., children being alone in their rooms while ‘attending’ the lesson); however,

on the other hand, some schools "considered live video lessons important for children's emotional and social health, and said these allowed teachers to spot potential safeguarding issues at home." (Ofsted, 2020b, p. 6). While the interviews in this research did not capture the reasons why these schools had implemented pre-recorded or live lessons (as these were ultimately up to the headteachers' discretion), the split in approaches even in this small sample exemplifies the heterogeneity in how schools dealt with remote teaching in the face of the pandemic. It is also important to highlight that the remote teaching situation during the COVID-19 pandemic was constantly changing, and as schools closed and reopened, or sent off 'bubbles' (i.e., groups with close contact) of self-isolating students, teachers had to constantly adapt. This was referenced, for example, by R4, who commented that she found the times when she was teaching some children online while also dealing with an in-person class stressful (Extract 6; line 1312), although in the same line she also indicated some relief in returning to her classroom.

Extract 6

Speaker	Line #	
R4	1311	Um, we went back to school then in July,
	1312	but that was just, and then a bit of a relief, but a little bit stressful as well teaching online.
	1313	Um, and then, and when we were in school we were teaching about groups of up to ten, so I
		was in this class and then I was streaming the lesson then to two or three other classes.
Interviewer	1316	Oh okay, yeah.
R4	1318	So we weren't really getting 1 to 1 time with anybody really, because
	1319	even
		though I was in the class with people, you were so, so trying to get all of
	1320	the technology and
		trying to stream, etc.

The circumstances in Extract 6 were reflected in the Ofsted briefing series, which stated that school leaders felt it was more challenging to offer remote education for small numbers of pupils given that teachers still had to teach in school simultaneously (Ofsted, 2020b, p. 6). To deal with this,

the report explained how schools would sometimes send out packs of work to children who were self-isolating, or set work on digital platforms as well as in exercise books. These situations are examples of how, during the COVID-19 pandemic, as schools had to adapt to consistently fluctuating scenarios, students experienced high levels of disruption to their lives, with knock-on effects on their educational development and emotional/physical health (Section 4.3.1). The following sub-section will explore in more detail how these respondents felt the move to remote learning had affected their literacy content in particular.

Returning to the new technologies implemented during COVID-19, other applications mentioned by participants included Seesaw, Quizlet, and Kahoot!. Seesaw, specifically mentioned by R1 (line 355) and R2 (Extract 7), is an application which is marketed as a "platform for student engagement" (Murphy, n.d.), where teachers can create activities to share with students, students can capture their learning in a portfolio, and families can see their children's work (based on settings that the teachers choose). A short report published by Seesaw evaluated the perceived efficacy of the application based on a survey of 300 teachers, and concluded that the tool is effective in engaging students and teaching them "21st century skills, like collaboration, problem solving, creative/critical thinking and digital citizenship" (Seesaw, n.d.). Furthermore, 76% of teachers reported that Seesaw saved them time and 92% reported that they had seen an increase in parent involvement since using Seesaw (of note, parent engagement was a critical theme that emerged in this study, and will be explored further in Section 4.2). However, it is important to highlight that the reports published by Seesaw were not peer-reviewed and lack a critical description of aims, methodology, dates, or author. However, in the present study, participants did discuss enjoying using Seesaw to supplement weekly Microsoft Teams meetings, and they elaborated on the way they used Seesaw to post similar activities to normal in-person teaching that the children would engage with (Extract 7). Some difficulties were encountered by R1 with the Seesaw app, mainly that

parents were able to contact her beyond normal working hours, which affected her work–life balance (line 355; this is discussed further in Section 4.3.2).

Extract 7

Speaker	Line #	
R2	450	It was lovely because we used Seesaw as
	451	an engagement tool. So I posted some activities online and the children would be able to do
	452	them and then post back and things like that. So we had a really good kind of interaction
	453	with children online and then very offline then. So once a week, it started off as every day,
	454	but then once a week we do a Teams meeting, like we're doing now. Which would be lovely
	455	too.

All participants described difficulties faced when attempting to manage the disruption caused by the sudden move to remote learning. Engagement was perceived as more difficult without the 'face-to-face' nature of regular teaching, and because they "weren't working with the children [them]selves" (R1; line 21), there were fewer opportunities to work 'on the fly' and adapt the lesson based on problems as they arose. In a 'normal' classroom, a teacher can monitor and observe their students to ensure everyone understands the work, but this is much more difficult remotely, and the interviewees described how it was difficult to not know whether their students were struggling if they did not ask for help. This was illuminated by R4 in Extract 8.

Extract 8

Speaker	Line #	
Interviewer	1353	How did you find that? You know, being online all day? You
	1354	know, because obviously you're used to teaching all day normally. Did it feel different?
R4	1356	Oh, gosh, yeah. Because the thing is, you haven't got any interaction, so you
	1357	couldn't gauge anything. Normally, in school, if you see that people aren't on task. You can
	1358	go by, you can help them, there's questions coming from them, so they sort of lead what you
	1359	do. Whereas online, you would just, well you were just hoping really that they were getting
	1360	it done. We were recording lessons. Some people were looking back at them afterwards. It
	1361	was that not knowing if they were doing it.
Interviewer	1363	Yeah. Not been able to keep track.
R4	1365	Yeah.

This difficulty in managing engagement and lack of creativity and spontaneity in remote teaching was echoed by other respondents. R5 attempted to maintain a similar lesson structure to in-person during remote learning, with lots of repetition and routine, but found it difficult and noted that she also missed the face-to-face aspects of teaching such as being able to observe children working and offer immediate feedback, as also described by other respondents. R5 mentioned how she had to maintain a necessary level of adaptability for remote teaching and modify lessons based on how she could keep the children engaged, as engagement levels differed between sets: "So my top set group were fantastic online. They were really, really engaged, really good. The [other] group were okay online. But, you know, they weren't they weren't as responsive." (R5; line 1971)

This shows that teachers had to adjust their teaching styles to combat engagement issues and continue achieving their educational goals, and one of the main methods to achieve this was to make their teaching style more prescriptive rather than loose and creative. Some described how they were expected to use different strategies to normal teaching, such as making the content more prescriptive and specific ("It's very, very prescriptive. And also, there's not a lot of room for

creativity. It's just much more when you are online, this is what you need to do and you need to do this now, kind of. Step by step instructions." (R2; line 556)), and this could be tricky as she could not get the kids as excited as she normally would in person. R5 again described being prescriptive because the children would struggle with open-ended or more creative tasks: "[…] if you gave them a very open-ended task, they found that incredibly difficult. Yeah, that's what we found anyway" (line 1962). Examples of prescriptive tasks included rewording sentences, anagrams, ordering 'jumbled-up' sentences (R1; line 694). Using different software was another strategy to increase engagement; R4 mentioned how they implemented different applications, such as Quizlet and Kahoot!, into their teaching to try and overcome some of these perceived engagement problems: "To be fair, a lot of work was done in school then trying to get interactive things put in. So then we started using mini whiteboards, Quizlet, Kahoot!, at least then you could see the type of things they were doing" (line 1339). However, no further comments were made on their perceived effectiveness. Reading Eggs was mentioned by R1 (line 400) as an application used to facilitate reading for her younger students, but again no comment was made on its perceived effectiveness.

To summarise these difficulties encountered during the move to remote teaching, R5 succinctly touched on the missing 'human touch' when teaching online: "Obviously teaching is a is a very people in a people-based profession. I mean, it's not our world to be down a camera." (R5; line 1749). This lack of in-person interaction led to some teachers feeling a sense of 'detachment' from their students, leading to anxiety over the children's welfare: "I didn't hear from some of the children at all in that whole time […] it does worry you when you don't hear from them in that long and you don't know how they're getting on." (R1; line 167). These sorts of anxieties will be further discussed in Chapter 4.7.

Despite many challenges, the adjustment to remote teaching was not seen as difficult by all respondents. For example, R5 commented that *"the school responded really quickly and we got into*

an online learning situation very fast" (line 1607), which she attributed to the fact that her school is an independent prep school where the parents pay *"quite a lot of money for fees"* (line 1606), and so there was pressure to ensure that the continued delivery of teaching was as seamless as possible (and, perhaps, the school had more resources to do so, but there is no evidence to support this). R4, who primarily taught older children, said that the transition to remote learning was not as difficult as expected. There was more planning involved to make sure all children were staying on task and able to keep up, especially if they had missed a lesson. However, as the children were older, they were more independent, and she described how they were easy to manage from the sense that the children had a different mindset as they were in 'big school' and working towards GCSEs (*"I think there's that mindset and that in secondary school to my gosh, I got lots of work to do. I've got to do it. You know, big school now." (R4; line 1553)*). Exams were tricky and involved collating evidence for grades rather than sitting exams as normal; however, grades worked out basically as expected. Exams were not discussed by the majority of teachers, as most were teaching primary years. What was discussed at length was the adaptations that had to be made to the curriculum, which will now be discussed.

4.6.2. Approach to Curriculum

Thus, as evidenced in the sub-section above, rapid changes had to made to 'how' teachers delivered their content throughout the pandemic; this sub-section will now cover the 'what', i.e., what content they delivered and how they felt the pandemic had impacted their curriculum delivery. This section will specifically focus on literacy content, as it is the interest of this thesis as a whole. The information in this sub-section was mainly gleaned from Questions 6–8 in the interview schedule and pertains to RQ1, 'How did teaching and learning practices change throughout the COVID-19 pandemic?'.

Respondents described having to make several adjustments to their curriculum content throughout the pandemic. As briefly touched on in Section 4.1.1, many felt they had to make their content more prescriptive, avoiding certain topics in favour of others, in order to ensure that children (and parents) could follow their instructions with ease without their in-person guidance. For example, their usual story writing topic was abandoned in favour of focusing solely on phonics for R2's class (Extract 10), and phonics was also mentioned as a key aspect of what R1 focused on (Extract 9), as well as writing- and formation-based work (i.e., handwriting).

Extract 9

Speaker	**Line #**	
R1	230	It was more just skills-based. So just phonics-based
	231	writing-based, formation-based. And that's all we could do for literacy because... They, they,
	232	I didn't want the parents doing the big pieces of work when they might not all do it and they
	233	might not all get those skills, then.

Extract 10

Speaker	**Line #**	
R2	684	But online,
	685	the expectation was to post different kinds of strategies. So instead of doing the kind of
	686	fluid, story planner or story mountain or story starter sentences and things like that, it was
	687	much more specific. So we needed to be teaching phonics more than anything else.

This focus on phonics teaching was echoed in a report by Ofsted conducted in October 2020 which surveyed teachers upon their return to face-to-face teaching (Ofsted, 2020b). The report states that upon their return to school after the first lockdown, almost all schools were concentrating "strongly" on phonics (p. 4), and that some schools were teaching phonics to all pupils twice a day or extending their phonics teaching into Year 3 (beyond where the curriculum would ordinarily stop teaching phonics). Of note, Extract 9 indicates that one of the key reasons R1 did not

want to offer creative tasks to the children was that parents would help the children too much to complete larger pieces of work, and thus the children would not gain the necessary skill levels in those areas (line 232); parental involvement is discussed further in Section 4.7. Examples of the specific, prescriptive tasks given in these interviews included rewording sentences (line 694), anagrams (line 695), and using SAT books which they would not ordinarily use in the classroom (R5; line 1959). R2 went on to illustrate how she found these tasks different from her usual practices: "Yeah, and you can read them stories when you're in the classroom, whereas you couldn't really online. So it was much more, much more prescriptive." (line 697). Some did attempt to read to their classes remotely, and R1 spoke of more success with this through pre-recording her class a story every day (line 250). R4, who taught older children, described how it was more difficult to teach GCSE English language and literature as it was trickier to annotate set texts or discuss literature to the same level of depth remotely. She adapted her curriculum to a 'flipped learning' approach (line 1444), wherein students prepared before they came to class. This worked well when children did prepare, but, as she remarked, if they had not prepared it was difficult to do the same lesson and to ensure the students caught up. Flipped learning has also been reported to be an effective instruction technique for remote teaching during the pandemic in university students (Rehman & Fatima, 2021). The importance of adapting a curriculum for online learning throughout the COVID-19 pandemic has been observed in other studies (Noor et al., 2020) and research has also highlighted that teaching methods that encourage active engagement, such as project or small group activities, mitigated the negative effects that the pandemic had on learning (Orlov et al., 2020).

Finally, one benefit of remote teaching noted by a respondent was the speed of work which could be covered, as, for example, R1 acknowledged that, to her surprise, they were getting through more work and topics than they normally would (Extract 11).

Extract 11

Speaker	Line #	
R1	17	Normally we have, in a week I plan for maybe even one maths, one
	18	literacy,
		one topic for the whole week. Um, but because the children were at
	19	home with their parents
		and it was just one on one with them. I found they were getting through
	20	a lot more as well.
		So I was finding that we were completing more and we were catching up
	21	a lot more than
		what we would have. That was positive.

Question 8 in the interview schedule asked the respondents whether they felt there would be any long-term impacts on their students after the pandemic. The predicted 'learning loss' due to COVID has been a widespread theme in academic and news articles over the last few years (Andrew et al., 2020; National Literacy Trust, 2020; Sibieta, 2021; Underwood et al., 2020), but qualitative evaluations of the long-term impacts of lockdowns on student achievement are lacking (Gore et al., 2021) and many reports concerning 'learning loss' only pertain to hypothesised or predicted loss (Donnelly & Patrinos, 2022).

Research is still being conducted into the effect of the pandemic on students' learning in the UK and worldwide. A recent systematic review (Donnelly & Patrinos, 2022), which analysed recorded learning loss evidence documented in studies conducted between March 2020 and March 2021, found that seven of the eight studies included in their review identified learning losses among at least some of their students. The small number of studies discussed was due to the novelty of this subject. These losses have been identified across a range of subjects, grade levels, languages, and geographic regions (including Belgium, Australia, the Netherlands, the USA, and Switzerland). The authors highlight that some students are experiencing more learning loss than others based on already present, intersectional inequalities (e.g., poverty, race), a fact that has been repeated in prior research (Education Policy Institute, 2020; National Literacy Trust, 2020). These intersectional

inequalities are echoed in the word gap ideologies which have been previously discussed in Section 2.1.

In this study, the interviewees also felt that there would be some long-term impacts for some children depending on how they had personally reacted to the pandemic and also on the support that they had received from their parents and the work they had done at home. These findings support the research on learning loss during the pandemic as discussed above. A brief summary of any perceived long-term impacts for each respondent is listed below:

- R1 felt that there would definitely be long-term impacts for some children (line 85), especially those who had not done anything at home, and she felt these children would be about six months behind where she expected them to be. She commented that their ability to catch up would be impacted by any future lockdowns.
- R2 said that some children would certainly catch up, but that it would take longer for others as they had "already settled into the behaviour of not doing things" (line 793). She felt that the ones who would take longer to catch up were those who had received more parental support.
- R3 described that a few of her children would definitely have their literacy impacted long-term, and she distinguished between those children who had spent all lockdown reading versus those who would only read at school (line 1136). For the latter, "their vocabulary has completely dropped, their ability to form coherent sentences has dropped. And it's then quite hard. It's really hard to then build that back up again to the level it should be." (line 1139).
- R4 was not sure, but felt that Year 7 may have been particularly affected as they missed their last year of primary school and some of their first year of secondary, where they normally get 'back to basics' in grammar and so the staff are not sure of all the children's

levels (line 1458). She also commented that there may be more long-term impacts on those children who had missed exams as they won't know what areas they have to work on (line 1462).

- R5 mused that the long-term impacts of the pandemic were really "the unknown" (line 2007), and commented that she liked to think her students would catch up academically. Emotionally, however, she felt that there would be a lot of long-term impacts on mental health, especially for older children (line 2020).

To assess the divide between underperforming and 'as expected' children, and determine the children's progress post-COVID, several schools had implemented assessments and interventions. R1 described how they had introduced afternoon intervention sessions led by teaching assistants (line 90) for those children who were not performing as well upon their return to school (assessed by a phase two phonics assessment), covering "some phonic work and some maths work", and she hoped this would lead to them catching up. In R5's school, children were assessed upon returning to school and those who tested below their chronological age were offered basic literacy support, but they were unable to offer individual support. Some children were identified as having specific speech and language difficulties after lockdown, and R5 noted that she thought these were especially common in children who were unable to read as much, those with complicated family lives, vulnerable children, and those with smaller families. Their school ended up with 6 or 7 speech and language referrals post-lockdown which was higher than normal. She described how their school implemented intervention programmes and that, by the summer of 2021, most children were making encouraging progress based on their normal spelling and reading assessments. Ofsted had also highlighted that, post-COVID, many schools were offering interventions to those children deemed to be struggling, implementing interventions such as one-to-one or small-group tuition, during school hours or sometimes before or after school (Ofsted, 2020b, p. 4). Specifically, Ofsted

highlights that interventions were often focused on reading, and sometimes mathematics and writing. To mitigate these long-term impacts, Ofsted offered schools additional funding, termed the "COVID-19 catch-up premium" (Ofsted, 2020b, p. 16).

In regard to specific literacy skills, the respondents observed several areas which they perceived to had been affected the most. For example, R3 described how some children had obviously spent a lot of lockdown reading, and so returned from lockdown with a high level of reading abilities, while others had done no reading and so had lost six months of progress. Interviewees felt that the main literacy skills that had been lost or not learned over lockdown were phonics (as discussed above) and writing. One respondent noted that the over-reliance on technology during remote learning had affected handwriting and spelling abilities, and thus they were spending more time on these skills now that the children had returned to school. Punctuation rules had been forgotten, children were using 'text speak' inappropriately (Extract 12), and this was attributed to the fact that students had been spending more time writing on devices rather than completing handwritten work during the pandemic. These devices contain features such as autocorrecting spelling, or the formats the children were writing in (e.g., text messages) did not require careful use of punctuation. These regressions were perceived to have not affected older children as much, as they were already more confident with writing (R3, line 1100). Another respondent suggested that the loss of writing skills was down to missing the basic teaching in primary school: "I think it's more to do with reading skills and basic grammar skills that I think they've missed time in primary school, especially Year 7 coming up. So I think [I'm] finding more basic errors in their writing." (R4, line 1450).

Extract 12

Speaker	Line #	
R3	1080	Um, I'd say that one of our main focuses when we came back was on reading
	1081	and spelling and stamina, overall, I think. Because a lot of the work they've been doing was
	1082	not handwritten. They're doing lots of stuff on devices, and so a lot of their actual writing
	1083	skills weren't as strong as they were. So yeah, handwriting and spelling in particular because
	1084	using devices which have like autocorrect.
Interviewer	1086	Autocorrect!
R3	1088	Exactly. I had so many children saying I just can't remember how to spell it. I
	1089	had one girl in particular who was who was always saying, "Look, it's just I'm so used to
	1090	autocorrect and typing and things like that". And a lot of sort of text-like writing the word
	1091	your as "ur" and things like that, was particularly common, in the beginning. And basic
	1092	punctuation and things like that. Missing commas and capital letters and full stops, which
	1093	again you tend to do more when you are messaging or typing, which then isn't reproduced
	1094	in that writing. That was one of the biggest challenges.

A focus on prioritising writing, particularly stamina and handwriting, upon return to school, was also reflected in a report by Ofsted carried out in October 2020, which stated that their surveyed teachers felt that “pupils had lost this [writing] ability during the first national lockdown, when they had not had the chance to practise their writing style and posture” (Ofsted, 2020b, p. 4). A study from Australia, which surveyed 310 schoolteachers, found that 72.6% of teachers felt that COVID-19 had affected writing instruction in their class, and this impact was mostly perceived as negative (Merga et al., 2021). The themes which emerged from their study showed that teachers felt that their students’ writing skills had been negatively impacted by interruptions and adjustments to the normal curriculum, leading teachers to focus solely on familiar tasks and skills; limitations to

instructional approach (e.g., less peer interaction, no in-person interactions to ensure children had learned the necessary skills); device limitations; and lack of or overreliance on parental support. These themes align with other research (National Strategy for Educational Research and Enquiry, 2021) and with the themes that emerged in the present study, especially in regard to curriculum adjustments, such as focusing on prescriptive strategies, which been previously discussed. Parental support and device limitations will be discussed further in Section 4.7. Reading does not seem to have been impacted as much, which we could ascribe to the stated continuation of reading work carried out by the teachers in this report (see below) and by parents at home, which has also been discussed as an important mitigating factor in children's continued literacy education in other research on reading practices in the pandemic (Chamberlain et al., 2020; Gore et al., 2021; Wheeler & Hill, 2021) and will be further explored in the next section.

On the flip side, some teachers were surprised by some of their children "surpassing expectations" (R2). R1 described how her top sets were doing well in general too (line 246) and noted that the imagination and vocabulary of several of her students was impressive, "especially the boys", who were using new words (e.g., *habitat, damp, soil*) that she had not explicitly taught them, and thus their vocabularies were better than expected (line 247). Extract 13 contains some further examples of how she had noticed their vocabularies improving. She attributed this to several potential factors: she had recorded them a story every day, maybe their parents were reading them stories, or maybe they were watching a lot of YouTube videos over the lockdown period.

Extract 13

Speaker	Line #	
R1	264	I definitely think it's improved. I mean, for
	265	example, one of the boys, we were writing our story and he was talking about a bug who's
	266	climbing up a mountain and then he saw a volcano and the volcano was erupting. You know,
	267	and he's Reception. And then we had some a little girl who was writing about an ice cream
	268	that was dropped. And she said she, the ice cream shop was closed, *sadly*. You know, it's
	269	like, where have you come up with these?

Some research suggests that this is not an isolated incidence of vocabularies improving throughout lockdown. Kartushina et al. (2022) evaluated the vocabularies of 1742 children aged 8–36 months across 13 countries and 12 languages, at the beginning and end of the first lockdown period from March to September 2020. They found that children gained more words than expected during lockdown, which could be due to the intense caregiver–child interaction during lockdown or the extra awareness that caregivers had regarding their child's development. Positive impacts of the pandemic on literacy have been observed in other research, as, for example, the National Literacy Trust highlighted that children were enjoying having more time to write on their own and think about their writing, and that they enjoyed writing on digital devices as well (C. Clark et al., 2020). A study conducted in New South Wales, Australia, found that the growth in student achievement during the 2020 school year only varied minimally compared to the 2019 year, based on assessments of reading and mathematics achievement (Gore et al., 2021). The authors posit that additional time spent reading at home, supported by family members, may have contributed to the minimal losses in reading achievement. All of this research highlights the important role of parental involvement in education during the pandemic, which will now be further examined.

4.6 The Role of External Support

With the shift to home learning, parents, who until then had played a smaller role in their child's education, were suddenly front and centre of the experience. Thus, alongside the impact of the pandemic on students and teachers, the ongoing crisis situation was also placing new demands on parents and caregivers. The impact of 'stay at home' orders meant that some parents were working from home while their children were learning from home, while other parents, especially those of key workers, were still working outside of the house. For parents who were working from home, they were suddenly expected to juggle their own work and ensure that their children could manage their schoolwork and keep up with remote learning. The respondents in this research described how they had much more contact with parents than before the pandemic and were having to manage parental worries and anxieties, but this also meant that they were able to observe which families were able to offer more support to their children. Thus, children from homes where the parents were furloughed, for example, and so who were able to be at home to support their children's remote learning, were perceived to perform better and return to school with the expected (or even better than) level of progress compared to those children from homes where parents were still working, or those from single-parent households or households with many siblings. Respondents acknowledged the differing circumstances amongst caregivers (R2, line 618) and remarked at length on the difficulties presented to them regarding caregivers. For example, the teachers had observed that many parents were struggling with managing their work and family alongside, for example, illnesses in the family or lack of devices for the household (Extract 14).

Extract 14

Speaker	Line #	
R3	1026	so I got to know some of the parents quite well.
	1027	And, and some parents were really struggling because of their own personal situations. With
	1028	everything, with family members being ill, with job insecurity, having suddenly all their
	1029	children at home and only having one device which they could use between all of them. I
	1030	think that was something that was a really big struggle for a lot of families.

This awareness of the challenging situation led to respondents feeling that part of their new role was managing parental anxieties and expectations, so much so that respondents felt they were 'also looking after' the parents as well as the students: "I kind of knew that you, you were guiding the parents and children through a difficult time. And they kind of did need you there, you know, if you know what I mean, they kind of needed to know that you were there at any time" (line 604). Respondents had to support parents through understanding how to help their children complete their schoolwork (R1, line 357; R2, line 633), and also had to manage parents' expectations (R2, line 674) surrounding their children's education. This sentiment has been echoed in the literature, which shows that parents had increased involvement in the minutiae of their children's education during the pandemic (e.g., ensuring differentiated instruction) and that parents also had higher levels of stress as they got to grips with their new role as 'parent/teacher' (Letzel et al., 2020). Furthermore, the Institute of Fiscal Studies found that many parents of both primary and secondary school students struggled with supporting home learning during the pandemic (Andrew et al., 2020), which was also noticed by respondents in this study. This increased pressure from parents alongside adapting to the remote education contributed to the decreased work–life balance discussed in Section 4.3.2, due to teachers receiving texts at all hours, for example (R2, line 622).

When asked about whether they had an idea of what separated the children who were underperforming to others, all respondents mentioned that they believed support at home was a key factor in how the children were coping emotionally and academically. Those children who had better support at home or had “stable homes” (R2, line 749) generally remained at the level expected of them and adjusted better to the online teaching, while those who the teachers believed were not being supported as well were underperforming. Respondents explained that children with parents who could spend a lot of time with them were performing at the level expected of them or even exceeding that level, while those children with busy parents who could not provide the same level of support were struggling. In this way, the success of the children’s education during the pandemic was determined in a large part by their parents. Extract 15 demonstrates this, as R1 compares the children whose parents did ‘everything’ with them to those who did ‘nothing’. The implication is that the latter were now behind and having to start from scratch when they returned to school, which meant that some students were further ahead than others.

Extract 15

Speaker	Line #	
R1	68	Some parents haven't done anything at all. So then obviously we, we come back, then we've
	69	got the children who've come on massively because the parents have been doing everything
	70	and more with them. And then we've got children who have literally done nothing. And then
	71	it's like starting from the start then.

Respondents specifically noted differences between children in supportive families where parents were furloughed and able to help their children with extra work versus those with working parents, single parents, or in large families where children had to help look after their siblings. One respondent commented that parents controlled a large portion of what her class learned during the pandemic, leading to a perceived learning loss amongst those with parents who did not encourage their children to complete their set work: “Parents have just kind of just, if they didn't want to do it,

they haven't. If they could, if the child could do it before, they can't do it now." (R1, line 24). R1 explained how she felt some children had come on more than what she would have expected normally, and felt that this was due to parents helping at home; she especially noted that it made a big difference if their parents were also teachers (line 49). This sentiment was echoed by R2, who explained that one child who surpassed expectations had a parent who was a teacher (line 711). R4 noted a "prominent gap" (line 1426) between the children who had been supported versus those who had not and also said that the children who had 'flourished' had done so due to structure at home (Extract 16).

Extract 16

Speaker	Line #	
R4	1392	we've come
	1393	across, you know, homes that are really, really, really supportive. And I think those are the
	1394	children that have really grown and flourished in lockdown because you can see that they've
	1395	had structure at home. Parents have been helping them with certain things. And I think it's
	1396	made that bond between home and school stronger. I guess it's not true
	[…]	for everybody.
	1427	Those that have had some sort of guidance at
	1428	home, and that have really slowed things down and done activities with the families, I think
	1429	they've benefited in some sense from lockdowns. I think
	1430	they've been reading more, I think they've had discussions around the tables. I think they've invested time really in, in helping,
	1431	whereas I really do see some students coming in and they haven't lifted a pen, they haven't
	1432	read anything other than social media posts etc.

Extract 16 also provides several other observations: the lockdown had impacted the bond between school and home in a positive way, children in families who spent more time together benefited from the lockdowns, and parent-led reading and helping made a perceived big difference in the children's development. Other respondents also discussed reading, explaining how some parents reported reading more with their children, while others did not get that same reading time;

these inconsistences then led to differences in literacy levels for some children. Respondents also explained that they felt that exposure to a variety of experiences during lockdown, such as quality time with parents (R2, line 716) and exposure to different types of media (R1, line 251), also contributed to reducing the learning loss during the pandemic for their students. The literature shows that caregiver-led home reading practices during the pandemic may have had a large contribution to mitigating the negative impact of lockdowns (Gore et al., 2021). One study found that parents read more often to their preschool children during COVID-19, although a statistical difference could not be observed (Wheeler & Hill, 2021). Furthermore, vocabulary development, which is significantly affected by caregiver–child interactions, improved during lockdown for one cohort of children evaluated during the pandemic, and this was attributed to the intense caregiver–child interactions during the pandemic as well as increased emphasis on child development (Kartushina et al., 2022). Merga et al. found that parent support (either too much or not enough) also had a major impact on writing development in their cohort of analysed students (2021).

However, some respondents expressed that they felt some parents were helping their children too much. For example, R1 described how she felt that one child, who was previously very independent and top of the class, had struggled during lockdown and upon return to school. The teacher felt this was due to her mother "spoon feeding" her the correct answers: *"[…]when you looked at the work that she'd sent back from online learning, it looked like she was still keeping up with it and doing all herself. But I think the mum had been doing it either for her or, you know, spoon feeding it to her. So she hasn't got, she lost all independence really with her work."* (R1, line 39). The second half of Extract 17 supports the idea that some children were really struggling with their return to school due to lacking their parents' support (line 807–8).

Extract 17

Speaker	Line #	
R2	795	But some of the children's reading has been very impacted. So I was doing
	796	key words with some of the children a couple weeks back and whereas the children that
	797	have had lots of input at home, they're flying. So some of them have done really, really well
	798	with the key words. But others that haven't had as much support are really behind in that
	799	respect. So I think that the children that are behind in an educational sense are very aware
	800	that they are. So that spikes an anxiety within them too.
Interviewer	802	Yeah, I bet.
R2	804	So whereas they didn't have that pressure of reading or doing some work or
	805	reading, writing or doing the number work or whatever, now they do have that pressure
	806	and they're really feeling it. So that, um, adjusting to not having to do it when they're at
	807	home, when mum or dad or can't do it and haven't got the support to,
	808	adjusting to now having to do it.

The results from the respondents in this study strongly reflect the qualitative and quantitative research which is emerging regarding the impact of caregivers during the pandemic. Caregiver educational involvement is a key contributing factor to the educational success of children (Epstein, 2018; Seginer, 2006). The impact of parental engagement with their children's learning is not a new area of research and has been found to have a generally positive effect on learning (Goodall, 2013, 2018; Jeynes, 2014). While traditionally it has been seen that education is the duty of the schools and wellbeing is the duty of the parents, now these duties are merging and the roles of teachers and parents are also merging (Levinthal et al., 2019). This merging of roles has only been hastened by the pandemic, as almost overnight boundaries and responsibilities were dissolved and new paradigms emerged. Suddenly, parents had to become much more involved in their children's

schooling. Research commissioned by the Welsh Government showed that during the pandemic, teachers valued engaging parents in the schooling process and parents themselves engaged with their children's learning in many ways during lockdowns, both through supporting formal learning tasks and prioritising non-academic learning experiences (e.g., play, cooking) (Goodall et al., 2021). This research also showed that parents appreciated direct and clear communication from schools and that a focus on supporting relationships with parents for wellbeing was emphasised.

Research commissioned by both Welsh and English governments has shown that the impacts of the pandemic were wide-ranging and that prior disadvantages and access to support at home were key factors influencing students' success during the move to remote learning (Chapman et al., 2021; Goodall et al., 2021; Spielman, 2020; Waters-Davies et al., 2021). The impact of the 'digital divide' and the exacerbating impact of the pandemic on vulnerable children has been a prominent theme in the literature (C. Baker et al., 2020; Coleman, 2021; Underwood et al., 2020), and also emerged in these interviews. The Institute of Fiscal Studies found that children from better-off families in the UK spent 30% more time on home learning than those from poorer families and had better access to resources for home learning (Andrew et al., 2020). Extract 18 presents an account of how one school tried to provide devices, but not all children received them and some just had to 'make do'.

Extract 18

Speaker	Line #	
R3	1236	The biggest
	1237	factor is social and economic situation at home. And how much help and support they've
	1238	had and also just physical devices they've been able to use during the lockdowns because
	1239	there have been so many. And if you, you know, you might have a parent that's really
	1240	supportive, but if they've only got one mobile phone to share between four children, that's,
	1241	that's not going to be great. So I'd say that's probably the biggest factor in terms of how
	1242	their literacy is coming back to school.
Interviewer	1244	And you said you work in quite a big school, so I guess they couldn't provide devices to children?
R2	1247	We had a few devices that we were able to get out to the ones extremely in
	1248	need. But yeah, there were some who just sort of had to make do.

However, it is important to highlight that some research showed that deprivation had a lower impact compared to support structures, i.e., children's experiences during the pandemic were affected more strongly by whether they had a good support structure around them and the amount of time they spent with families/carers (Spielman, 2020). These findings were echoed by this research, as all teachers described that they felt the students who had better support at home coped better with the changes and were able to better handle continuing their schoolwork.

4.7 Impact on Wellbeing

This section will now turn from the educational impacts of the pandemic to the physical, emotional, and mental impacts mentioned by the respondents. Although wellbeing was not a specific part of the interview schedule, the answers to several of the first interview questions revolved around mental health, workload, and wellbeing, and so this became an important theme of

this research. The mental health impacts of the pandemic have been widespread (Mind, 2021), with some calling the mental health crisis a 'second pandemic' (Choi et al., 2020). These impacts derived from the extreme disruption that the pandemic caused to people's normal routines, isolation, loneliness, as well as disrupted health services and difficulties accessing treatment. The interviews reflect this; all respondents discussed the wellbeing of staff and students at many points throughout the interviews, especially when asked to summarise the general themes of the past year. Student wellbeing will be examined first, followed by staff wellbeing.

4.6.3. Student Wellbeing and Mental Health

Mental health impacts throughout the pandemic have not been limited to adults, although few peer-reviewed and non-peer-reviewed reports have focused on children and young people under the age of 18 (Ford et al., 2021). A report from the mental health charity Mind found that 68% of the 1756 young people aged 13–24 surveyed experienced worse mental health during the pandemic. Similar findings were echoed by England's Mental Health of Children and Young People survey, which reported that there was an increase in probable mental health problems affecting those aged 5–16 years old in England (from 10.8% in 2017 to 16.0% in 2020), with those surveyed citing sleep disturbances, loneliness, disrupted access to healthcare and physical distance from adults outside their family contributing to decreased mental health during the pandemic (Newlove-Delgado et al., 2021). Furthermore, a rise in referrals to mental health and eating disorder services, as well as further effects on adolescent health, have been noted (Reed & Ort, 2022). In these interviews, mental health problems were noticed by teachers and were attributed to the transition to remote learning and then coming back to school and the direct impact of COVID on children's health, their families' health, and the group isolation/'bubbles' systems.

All respondents mentioned concerns for their pupils' wellbeing throughout the pandemic and several noted that their pupils seemed to be suffering from increased levels of stress. In the

younger years, this was apparent in observed regression (e.g., in social skills, separation anxiety, and in toileting behaviour). For example, one respondent who teaches nursery-aged children noticed that more children were struggling to come to school compared to pre-pandemic: *"We've got children who come in crying, still in reception, and they didn't come in crying before lockdown and now they're coming in crying every day because they miss their mum. We have that quite a lot, 'I miss my mummy.'"* (R1, line 196). Some were struggling more with their emotions and confidence, e.g., not wanting to take risks, feeling insecurity upon their return to their classroom, and expressing that they didn't feel they could do the work (R1, line 720). In the older years, mental health impacts were observed in increased observed levels of anxiety, depression, and stress: *"We found a lot more children have been more anxious, I'd definitely say that. Anxiety is definitely something that we've picked up more of."* (R5, line 2039). These reactions were being observed more clearly upon the children's return to school. Extract 19 summarises a few of the things that one respondent had noticed in her class: children dealing with deaths in the family, children struggling with the transitions between home/school, an awareness amongst the children of the pressures of the pandemic, dealing with new hygiene routines, and being overwhelmed in the sensory-heavy classroom environments upon return to school when they had been used to being at home, and again the perceived increase in anxiety amongst the students.

Extract 19

Speaker	Line #	
R3	1156	Well, I had a few children lose family members, so obviously that was, having
	1157	to manage that. A lot of children have developed anxiety.
Interviewer	1159	Yeah?
R3		And there were some who were really struggling to come to school, not wanting to leave their families, not wanting to come out. A lot of them are really excited to see their friends again. But then again, that transition just to the freedom you have at home and the rules of school, and especially all the new rules that were in place, like washing our hands sort of five or six times a day and of the timetable not being able to go and see, a lot of, often the year six's get lots of roles of responsibility throughout the rest of the school, like going to read to children or things like that. So they weren't able to do anything like that. And, um, that... I can't remember what I was saying.
Interviewer	1170	We're just talking about well-being. When you said, you said a lot were anxious to come back.
R3	1173	Lots of anxiety about talking in class as well. And getting very overwhelmed by sort of sensory things, like lots of noise suddenly, being in a whole class of children, being overwhelmed by the amount of work and the pressure and expectation on them to get that work done to a certain level because those expectations sort of filter down.

The ‘pressures’ mentioned in Extract 19 that the children were experiencing were echoed by other respondents, leading to resistance to being in a classroom environment and a struggle to be independent again: *“[…]we're back from Easter now for six weeks and still we're having, from some of the children, still we're having that resistance. To kind of settle and concentrate, and engage.”* (R2, line 808).

Isolation and lack of socialisation during the pandemic were discussed by participants in this study. Respondents noted the impact of lack of socialisation, describing how some children were coming back to school far more isolated and apprehensive about independent learning (R4, line 1389), and mentioning that some children had had "a lot of time to themselves" (R1, line 331), and so were having to adjust to being back in a classroom with many other people. On the positive side, respondents also discussed how some children were really happy to be back in the classroom and amongst their friends again (R2, line 470; R3, line 1163). There was a sense among respondents that they were having to 'reteach' basic rules and re-establish boundaries in the classroom once the children were returning (R2, line 533). This extended to the curriculum, too. Respondents mentioned that now that they were back into school, they found they were recapping or going back over the same things more than usual because the children were not confident enough to move onto other topics *("[…] we've had to go back. We're recapping things more than I have before*." (R1; line 207)).

Research has shown that mental health deterioration was worse during the pandemic in children whose families were already struggling (Ford et al., 2021) and this was also mentioned by the respondents: *"Children that have got [problems] in their home life, those are the ones that struggle. […] There is [sic] different pressures on parents. Ergo, their home life, the work life and some children have adjusted very well and other children haven't, you know."* (R2, line 751/780). This is indicative of the present inequalities that were only exacerbated during the pandemic, leading to worse mental health outcomes for vulnerable and disadvantaged children due to varying factors such as poverty, difficulty accessing equipment (Mind, 2021).

In response to the mental health impacts, some schools had made wellbeing a particular focus. One respondent acknowledged that her students had changed a lot over lockdown, and upon returning to school they had needed more flexibility, which resulted in them discarding their strict uniform policy (R5, line 1989). They also extended their wellbeing support services in the light of the

pandemic: R5, who works in an independent school, described the external counselling service offered by a children's mental health charity (Place2Be) they had partnered with and spoke at length on the success of the project: "Having that support in school has been really valuable. Because what that service allows to happen is for a child to have an absolute ringfenced period of time in the week if there is 1-to-1 counselling when they know that they're going to get dedicated time." (R5, line 2041).

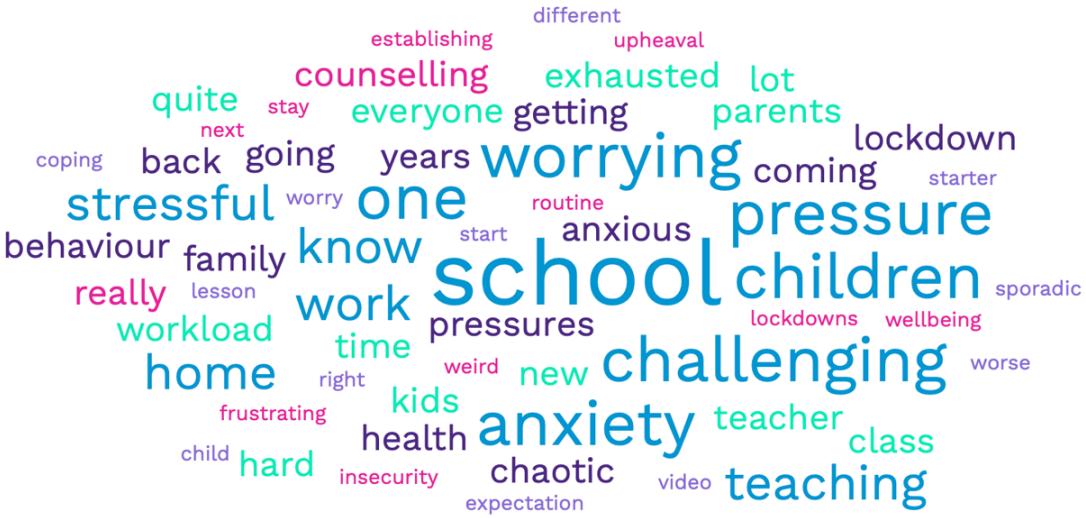
A non-peer-reviewed cost–benefit analysis conducted on behalf of Place2Be found that the individual counselling services that they offer as part of their 'whole school' interventions have a positive impact on the mental health of young children and have a favourable cost–benefit ratio (Gomez, 2022). While this is a really positive example of how schools can help the wellbeing of their students, it is important to note that the R5's independent and fee-funded school paid for this service and did not ask parents to contribute anything towards it; this sort of service would not be accessible for other schools or parents, especially those in lower-income areas. Finally, just as research has shown that lack of external wellbeing monitoring impacted children's mental health in the pandemic, (Newlove-Delgado et al., 2021), two respondents (R4, line 1377; R5, line 1674) specifically highlighted safeguarding concerns, explaining that the pandemic and remote learning made it more difficult for them to ensure that their pupils were not experiencing issues at home; another interviewee also spoke of the worry that came if they did not hear from their students for a while over lockdown (R1, line 168).

4.6.4. Staff Wellbeing and Workload

When discussing their own wellbeing over the COVID period, some of the words used to describe the year included: *stressful*, *weird*, *worrying*, *chaotic*, *challenging*, and *exhausting*. Figure 4.2 shows a word cloud of the most common words used in the parts of the transcripts coded for 'anxiety/stress'.

Figure 4.2

Word Cloud of Responses Coded with 'Stress/Anxiety'



These feelings of stress and uncertainty stemmed from several key areas: changing working practices (i.e., the move to online learning, adjusting curricula, etc.), worry for their students, health concerns, increased workload, and the changes to work–life balance. COVID-19-specific issues were also raised, such as dealing with colleagues having to isolate, becoming sick from COVID-19 themselves, and, for one respondent, dealing with the death of a colleague due to COVID-19. A general sense of uncertainty was illustrated by several respondents and Extract 20 shows how the respondents were in a constant state of flux as they waited to see what would happen with the lockdown situation in their areas. It also illustrates other feelings felt by the respondents, e.g., that the year was chaotic and a large upheaval.

Extract 20

Speaker	Line #	
Interviewer	490	If you could summarise the last year in just a few words, maybe
	491	three words, what words would you choose to summarise it?
R2	493	I think... From an education perspective?
Interviewer	495	Just from personal or educational.
R2	497	Chaotic.
Interviewer	499	Yeah.
R2	501	Upheaval.
Interviewer	503	Yeah. That's a good one.
R2	505	Insecure. Insecurity, I guess.
Interviewer	507	What do you mean by that?
R2	509	I just mean that you don't know... If it's actually... Outside of the lockdown,
	510	you don't know if you're actually gonna stay there. You know, it's just this insecurity of not
	511	knowing what's coming next. Or, you know, we're back in school, are we gonna stay, or do
	512	we have to isolate again, or will there be another case? And, you know, it was the insecurity
	513	of establishing something and not knowing if you're going to see it through.

A main factor that negatively affected the respondents' own wellbeing was increased workload. Working hours had increased during the move to remote learning due to increased demands for creating new educational resources or adapting prior resources, as described in Section 4.6. Work–life balance was brought up as all respondents had worked from home at least part of the time, and some found this adjustment difficult, especially if they did not have a specific area from which to work or were working in the same household as their family. Remote working presented challenges in staying connected with their colleagues, and respondents felt they had to "work that much harder keeping in touch" (R5, line 1632). It was also more difficult to develop a sense of connection with their students: "trying to keep the relationships going with the children over Zoom was not impossible, you know, it was manageable but, but it's just a bit different" (R5, line 2164). Increased demands from parents were raised as issues, with respondents explaining how parents would be messaging them at all times during the day/night and at the weekend. This led to a feeling

of being unable to ‘switch off’ from work (Extract 21 and R3, line 1007) as they were often responding to parental requests on their phones until late at night (R2, line 598). Specifically, R2 said she often would not be able to shut down her computer until 9pm in the evening (line 603).

Extract 21

Speaker	Line #	
R1	364	When you're working from home, I kind of felt like I needed to see what it was straight
	365	away. Like, I couldn't switch off from the notification that it was there.
Interviewer	367	Yeah.
R1	369	And then you can't switch off then. You're constantly thinking about work.

These interview results reflect research on teacher wellbeing. The increased workload and blurring of work–life boundaries was shown in these interviews and reflects findings from the literature which show that teachers’ wellbeing was affected by stressors arising from the pandemic situation, such as changing teaching practices and working from home (Jakubowski & Sitko-Dominik, 2021). Research has also shown that work-related anxiety increased in the week before lockdown and again in the week that schools reopened in June 2020, based on longitudinal survey data of around 8,000 teachers (Allen et al., 2020). This anxiety spike was more pronounced in headteachers, who had higher and more sharply increasing anxiety levels throughout lockdown. One in five headteachers also said that the experience during COVID made it more likely they would leave the profession (compared to one in ten teachers). Their research did show, perhaps unexpectedly, that the anxiety levels of teachers in general (not headteachers) returned to pre-COVID levels after the end of March 2020. Their results also showed that teachers who were part of their school’s skeleton staff, i.e., those teaching children of key workers in person, had higher anxiety levels at the start of the pandemic compared to those working from home. This is in agreement with their findings that providing ‘live’ online teaching was the most stressful teaching activity during the move to online

teaching and learning (compared to delivering pre-recorded lessons or nothing). This was in part reflected in these interviews as well, in that some teachers struggled streaming live to their students (Section 4.6). Using broader measures of wellbeing, Allen et al. (2020) found that teachers were more likely to agree that the pandemic had negatively impacted their mental health, although overall levels of wellbeing did not change between October 2019 and April 2020.

These mixed findings support some of what came out in these interviews, as some teachers said that they struggled with the constant changes and challenges of the year, while others said that they coped better than they expected. Thus, although the pandemic had obvious negative impacts on staff and students and presented several challenges, some positive aspects emerged from the interview data related to wellbeing and mental health. For example, although increased workload was mentioned by all, some respondents also acknowledged that, at times, they enjoyed working from home and found that they were able to catch up more and get more done. R1 said that she found the job less stressful than normal (line 9), and R4 described the entire experience as a challenge, but one that was “quite positive” (line 1525), in that she learned a lot over the course of the pandemic. R5 remarked that she was glad to see how her school had realised hybrid working was possible (line 2147). R2 also said that she still enjoyed working from home (line 595) despite the challenges presented in Extract 20.

Thus, a key positive aspect that came out of these discussions was that it highlighted the adaptability and resilience of both staff and students. All respondents remarked that, although students did struggle in some ways, in others they demonstrated remarkable flexibility and willingness to overcome challenges and try new things. R5 described how she felt the school and staff “coped really well” (line 1624), managing to host extracurriculars, encourage music and creative pursuits at home, and even put on a school play remotely. She also acknowledged how their school was able to be flexible and adapt to the needs of the children, such as by, for example,

dedicating afternoon sessions to sports or creative topics, rather than 'heavy' subjects such as French (which some "kids couldn't handle" (line 1954)). Finally, an increased sense of 'connectedness' was noted by all respondents, between their colleagues and their students. There was a sense that they had 'banded together', and several acknowledged that maintaining contact with their colleagues was vital during the pandemic. This feeling of 'togetherness' and social support is one key component of preventing burnout amongst teachers, especially in times of increased job demands (Brouwers et al., 2011; Russell et al., 1987). The job changes experienced by the teachers interviewed in this research involved the already discussed teaching and curriculum changes, as well as working from home, but also changes to how they approached their day-to-day tasks, such as ensuring more stringent standards of hygiene than normal (R1, line 186).

It was also acknowledged by respondents that when they were able to return to school, it was special to see their class again, and that her children were generally very happy to be back in school with a 'normal' structure and flow, and that it "means more now" (line 455) to her to see her children happy and safe after the chaotic period. This was echoed by R5, who stated that what made teaching worthwhile was being able to come back after the difficult adjustment periods and see her students' "smiling faces" (line 2162).

4.8 Limitations

This chapter has presented a myriad of qualitative evidence, most of which aligns with the increasing research on the impacts of the pandemic on teaching and learning in the UK and worldwide. However, there are some limitations to the work carried out in this chapter. First and foremost, the impact of the pandemic on education is a highly complex topic and the faceted nature of the problem presents several issues when attempting to measure it. The measurable impacts on children are confounded by associated and overlapping factors, both external (such as

socioeconomic status, geographic location, and race) and internal (such as inner resilience and individual differences). Furthermore, unless we have pre-pandemic control data, it is even more difficult to quantitatively measure the impact of the pandemic, which is a problem that has been highlighted in recent research (Donnelly & Patrinos, 2022).

Importantly, this experiment was not a planned part of the current thesis project and, although the novelty of the research is a strength of this study, the designed interview schedule, while crafted with care following procedures outlined in the literature, was not piloted. Furthermore, the sample size was small at five teachers across four schools in England and Wales only, lacking participants from Scotland or Northern Ireland. The four nations were not considered critically throughout the study, with examples and data mostly focusing on England and Wales; this is due in part to the inherent bias for research to focus on England, and in part due to time, focus, and length constraints of the thesis (i.e., systematically presenting data for all four nations would increase the length of the chapter, and is also beyond the scope of the research questions).

Even in light of these limitations, qualitative data in educational research can make a valuable contribution to causal inquiry and be a rich and valuable source of what works in education (Maxwell, 2012). Thus, although we have no statistical evidence and we acknowledge the limitations of presenting decontextualised quotes as data (despite attempts to acknowledge the contextual factors surrounding these data) (Talmy, 2010), I posit that the observations presented in these interviews are evidence of the causal impacts of the pandemic on children and teachers in the UK.

4.9 Conclusions

There remains a wealth of cross-cutting research, policy, and government interventions needed to address the impacts of COVID-19 on education in the UK. This research discusses the experiences of several schoolteachers in England and Wales and highlights the key impacts of the

pandemic on both staff and students, and is an original contribution to the emerging research of the impact of the pandemic on both teachers and learners in the UK. The data presented in this study found that teaching content and styles drastically changed, staff and students both had to adapt rapidly to a constantly changing situation, and several challenges emerged during this transition. Staff workload increased as the line between work and home was blurred, and parents became an increasingly influential factor in the educational success of their children (according to those interviewed). Some students struggled and staff noticed changes in emotional and mental wellbeing as well as impacts on grades and schoolwork. However, respondents in this research also brought up positive aspects of the pandemic, such as adaptability, resilience, and increased connectedness to pupils and parents. This resilience is not to be forgotten: as research moves forward, it is important that, when we acknowledge the impact of COVID-19 on young people, we do so while also acknowledging the strength shown by children and teachers across the UK (Senedd, 2021). The strength shown by all throughout this crisis can be succinctly summarised by R5: “I think we've all found that, that we can adapt. Yeah. You know, that we, we are adaptable and we are resilient, really” (line 2150).

5. A Lexical Investigation into the Oxford Children's Corpus of Writing

This chapter presents an investigation into a corpus of short stories written by children for a national writing competition in the UK. These samples are held in the Oxford Children's Corpus, a large corpus of writing by children in English collected by Oxford University Press. This repository of original children's writing has received some attention in the literature (Banerji et al., 2013; Hsiao et al., 2021; Wild et al., 2013) and this chapter builds on this research to contribute to the literature on aspects of children's writing in English. This study employs a quantitative corpus linguistic (QCL) approach to quantifying specific aspects of written language use in a sample of children's writing. This is a useful method for researching writing development and illuminating developmental trends in language use.

Sketch Engine is a corpus analysis tool where users can create and analyse bespoke or pre-loaded corpora to identify linguistic features (e.g., collocations) in context. Sketch Engine has a variety of ready-to-use corpora in several languages and a number of tools to analyse a corpus. Established by Lexical Computing Ltd and Adam Kilgarriff in 2004, Sketch Engine is able to execute a number of useful functions for corpus linguists, including word sketches, collocations, and concordances. Sketch Engine can also provide frequency data, keyword analysis, and can show collocations and concordances for single- or multi-word search terms. Sketch Engine is a particularly useful tool for lexicographers who wish to explore patterns of language use, and this has led to it being used heavily by dictionary publishers and university presses (Kilgarriff et al., 2014). Sketch Engine also includes parallel corpora for translation and corpora from specific time periods so that linguists can study language development and change.

In this research, it is used to lexically analyse the entire Oxford Children's Corpus, presenting frequency data for different parts of speech and keywords. A handful of keywords are analysed using Sketch Engine's word sketch and concordancing tools to investigate the contexts in

which children use these words in their own writing. Then, a smaller corpus of the winners' stories from the last three years is constructed to facilitate more fine-grained offline analyses into lexical diversity and sophistication. Thus, using comparative corpora and Sketch Engine, this chapter highlights some differences between writing by children, writing for children, and writing by and for adults, and illuminates patterns in children's language use in one particular genre (creative writing) in one context (submitting to a competition).

Quantitative corpus analyses of children's writing generally focus on assessing how writing changes over time, or on what distinguishes 'successful' writing from the rest. This research thus operates along two axes—time and quality (Durrant, 2022a)—and it is along these axes that this chapter is structured. In the discussion, the results and the competition/corpus itself will be discussed under the language policy framework encapsulating this thesis. A literature review follows this introduction, followed by two studies (the larger-scale Sketch Engine analysis, and the smaller offline analysis), a discussion, and an analysis of the limitations of this corpus and areas for future work.

5.1 Literature Review

A corpus is a collection of texts which can be examined for specific language features. Sophisticated computer software and developments in corpus linguistics and natural language processing have enabled faster and more reliable descriptions of changes and patterns in texts (Durrant et al., 2021). Quantitative corpus linguistic (QCL) approaches enable the analysis of large amounts of data, and these larger sample sizes allow researchers to find patterns in language use that may not be otherwise readily observable (Sinclair, 1991). This is important for studies of first language users, as research is increasingly highlighting that school-age development is not marked

by the emergence of new syntactic structures but rather by writers' abilities to handle increased complexity and genre differences in their writing (Applebee, 2000; Bazerman et al., 2018).

Finding the linguistic features of a text is one thing, but extrapolating to interpretations about communicative aspects is another. QCL approaches are based on samples of actual written use produced in specific communicative contexts, capturing writing proficiency 'in situ' (Durrant et al., 2021). However, this means that QCL only considers writing as a product, ignoring the inherent psychological, sociocultural and physical complexities involved in the act of writing and which have been touched on in the literature review (Bazerman et al., 2009, 2017; Grabe & Robert, 1996; Herrington & Curtis, 2000). Thus, corpus methodologies only offer one piece of the puzzle and bearing these limitations in mind is key when implementing QCL approaches.

Research quantifying writing development generally focuses on the areas of vocabulary, syntax, formulaic language, and cohesion. A systematic review of the literature from 1945–2015 identified 104 studies on L1 language development in the writing of children in Anglophone countries, distributed in the following way (with the caveat that some studies looked at multiple areas, and so these numbers will not add up to 104): 66 on syntax, 45 on vocabulary, 19 on cohesion, and 0 on formulaic language (Durrant et al., 2021). Vocabulary research has included evaluating lexical diversity (Malvern et al., 2004; Malvern & Richards, 2012; McKee et al., 2000b) or using frequency-based tools to quantify coverage of high-, medium- and low-frequency words (i.e., lexical sophistication) (Durrant & Brenchley, 2019b; Kyle et al., 2018, 2020; Kyle & Crossley, 2015). These variables are often compared against subjective ratings or evaluations to investigate relationships with quality, or development is tracked diachronically to evaluate correlations with age (i.e., the axes of quality and time).

This chapter will focus solely on vocabulary. Lexical richness measures (Read, 2000), introduced in Section 2.4, in L1 English-speaking children generally increase as children age, along

with associated increases in perceived quality. For example, lexical diversity of writing tends to increase as children get older and more lexically diverse texts are also assessed as higher quality (R. A. Berman & Nir-Sagiv, 2004; S. A. Crossley et al., 2011; Malvern et al., 2004; Olinghouse & Wilson, 2013; Uccelli et al., 2013). Lexical sophistication also increases, as children use more low-frequency words and more abstract words as they mature, as well as more 'academic' vocabulary, and low-frequency vocabulary is associated with text quality (Malvern et al., 2004; Durrant & Brenchley, 2019). However, there is no evidence for a relationship between lexical density and development (Johansson, 2008; Uccelli et al., 2013). Children also develop an increased awareness of register and style as they age. Genre impacts vocabulary richness measures, with, for example, narrative texts giving higher MTLD values than informative or persuasive texts (Olinghouse & Wilson, 2013).

There are several important methodological considerations that must be addressed before we begin research into lexical richness. First is that genre and contextual factors must be given their due consideration. Little research has looked at patterns of language use and how these patterns develop or are influenced by specific genres, topics, contexts, and audiences (Durrant et al., 2021). More research on patterns of language use in specific contexts will illuminate what these patterns look like in different texts and discover what drives quantitative patterns of language use. In the UK, the mastery of writing across genre is embedded in the English curriculum (Department for Education, 2014). Writing in different genres involves writing with a particular communicative context in mind, whether this be making a persuasive argument, writing a letter, or developing a narrative. This is a complex task that must be carried out considering the audience and overall purpose of a text. The idea of different communicative contexts is key when analysing writing development, as the characteristics of writing will differ depending on the communicative context, or genre, at hand (Grabe & Robert, 1996). It is, surprising, then, that research on writing often

overlooks or only partially reports on genre, but this may be in part due to the inherent complexity involved in controlling for it (Durrant et al., 2021).

After genre and contextual factors, it is also important to clarify the operationalisation of linguistic measurements, as lack of clarity leads to the inability to examine whether different studies have measured the same thing. Transparency about what is being measured, clear definitions of the construct under examination, as well as critical and active discussions of the hypothesised constructs are vital for future linguistic assessments of writing. Furthermore, measures must be linked to one another, and consideration must be given to the full scope of the target construct, as no linguistic construct exists in a vacuum. For example, lexical sophistication cannot be separated from lexical diversity, as is often done in writing research, because they interact with each other (Durrant & Brenchley, 2021). Other factors such as collocations or syntactic variables may be worth further investigation alongside vocabulary measures. Research on vocabulary would benefit from more fine-grained measures, as generalised mean word frequency hides an underlying complexity in vocabulary use which is revealed when vocabulary is examined per separate parts of speech. Integrating simple grammatical constructs into vocabulary examinations allows for highlighting patterns that would not be seen if these categories were not differentiated.

An example of the links between different constructs is that frequency has been found to interact with parts-of-speech, lexical diversity, and register and models of lexical sophistication as low-frequency, register-appropriate words (Read, 2000) may not be the most appropriate for studies of children's writing (Durrant & Brenchley, 2019b), and more work is needed in these areas to uncover the connections between different lexical measures. An interaction between frequency and PoS has been reflected in research on children's writing: younger children's writing is characterised by repetition of high-frequency verbs and adjectives but low-frequency nouns (due to the use of nouns that do not occur in general adult-directed writing, e.g., *fairy*) (Durrant & Brenchley, 2019).

This highlights the importance of measuring a range of variables in order to view a complete picture of children's writing. It also shows how something as seemingly straightforward as frequency needs careful consideration and operationalisation to ensure that the gathered results are meaningful.

Another example of this interaction between frequency and parts of speech can be found in regard to children's reading materials as well as their own writing. Quantitative corpus linguistic approaches have shown that writing for children differs from writing for adults in several key ways (Wild et al., 2013). Comparing a corpus containing child-directed fictional texts (the Oxford Children's Corpus, which is described in more detail later in this section) to a generalised adult-directed English corpus (the Oxford English Corpus) showed that fictional texts for children tend to have specific features: a large focus on the natural world (living creatures, plants, food), objects, communication (through direct speech), with an emphasis on size, speed, bravery, and fear, as well as physical space. Adult fiction, when analysed from a corpus linguistic perspective, is more concerned with time, intimacy, law and order, society, and beliefs (Wild et al., 2013). There are, however, remarkable similarities between fiction written for adults and fiction written for children in terms of frequency data for words and parts of speech. Furthermore, this study showed that the child-directed corpus contained more modals and auxiliaries, more nouns and pronouns, and overall shorter words than the generalised corpus. This suggests that there is core vocabulary consistent with narrative texts written for children and this core vocabulary differs when compared to other corpora. This finding was reflected through analysis of the CLIIP corpus by Thompson and Sealey (2007), which found that while narrative fiction written for both adults and children is characterised by some similar linguistic features, writing for children does differ in ways mainly related to how children perceive the world. The language in children's books has also been found to contain less frequent words and be more lexically diverse and denser compared to child-directed speech (Dawson et al., 2021). This reflects research showing that written language is more lexically diverse

and syntactically complex than spoken language (Malvern et al., 2004). Thus, written language is a rich source of vocabulary for children and being able to compare children's reading input with their writing output is meaningful.

A final thing to note is that studies often show variability in writing samples received, which are samples generally collected in controlled conditions to create a corpus with a specific aim in mind. The variability is due to the different experimental procedures implemented in different studies and general lack of operationalisation of constructs (Durrant, 2022a). Thus, the analysis of larger corpora of texts, especially those collected under naturalistic conditions, may allow for deeper investigation into the lexical features of children's writing and provide robust evidence for the study of the development of children's L1 writing.

Based on the identified need to operationalise linguistic assessments of writing quality more extensively, especially in younger writers, as well as consider previously neglected issues such as genre, this investigation used the Oxford Children's Corpus to conduct several investigations. The Oxford Children's Corpus, held by Oxford University Press and hosted on Sketch Engine, is split into two parts: a reading corpus, including fictional books and materials children are likely to encounter, and a writing corpus, which includes writing produced by children aged 5–13 for the annual BBC Radio 2 500 Words Writing Competition. The former has been investigated in regard to relative clauses and in a comparison to an adult corpora (Hsiao et al., 2022; Wild et al., 2013) and the latter corpus has been investigated in one study on gender bias, which found a clear bias for male names in the corpus (as characters), especially in regard to the boys' writing (while girls were more balanced when choosing male or female characters) (Hsiao et al., 2021). This was linked to the gender bias that is present in literature on the whole (for both children and adults), with the authors concluding that the literature children consume influences their own writing (Hsiao et al., 2021).

Further analysis, contextualised in the methodological considerations discussed above, of this large corpus of children's writing is worthwhile.

This chapter sought to answer the following questions:

1. What characterises the writing that children submit to the 500 Words short-story writing competition?
2. What characterises the writing of the winners of the short-story competition? To what extent do the winners' stories differ from the other entries?
3. To what extent does the writing of the 5–9-year-old group differ from the writing of the 10–13-year-olds group?

These questions reflect the need to look at patterns of language use in context to see what drives quantitative patterns and illustrate what these patterns look like in text. They also allow examining these patterns in relation to assessments of quality and age (i.e., the two axes defined by Durrant (2022)).

To answer these questions, two experimental studies were designed. The first aimed to analyse the entire corpus of the BBC 500 Words short-story writing competition using Sketch Engine, where it is stored and managed by Oxford University Press. The second involved the creation of a 'mini-corpus' of the winners' stories in order to perform a complementary and more detailed lexical analysis and relate lexical features with assessments of perceived quality. This winners' corpus was analysed separately from Sketch Engine using different tools and software for measures of lexical diversity, density, and syntactical complexity. It was also uploaded to Sketch Engine for keyword and other comparisons.

5.2 Study 1: Sketch Engine

This first section attempts to answer the first research question presented above: *(1) What characterises the writing that children submit to the BBC 500 Words short-story writing competition?* To answer this question, lexical analyses of the BBC 500 Words corpora were conducted using Sketch Engine. In this section, the corpus under investigation will first be presented, then Sketch Engine's tools and functions will be discussed, the methodology will be summarised, and then the results will be presented.

5.2.1 The Oxford Children's Corpus

The Oxford Children's Corpus in its current state exists in two separate parts: the Oxford Children's Corpus of Reading (OCCR) and the Oxford Children's Corpus of Writing (OCCW). Both are annotated corpora tagged for PoS and some metadata which vary depending on the corpus. The former was identified as "the first of its kind" in 2013 (Wild et al., 2013, p. 190). The OCCR contains over 60 million words of writing targeted at 5–14-year-olds. It includes fiction, non-fiction, magazines, curriculum materials, and texts from the internet (e.g., the Minecraft Wiki) from the 19th through to the 21st century. It is used by lexicographers to develop children's dictionaries (e.g., the Oxford Children's Dictionaries) and track trends in children's language (Banerji et al., 2013) and has also been used in research of children's literature (Wild et al., 2013).

The second part of the Oxford Children's Corpus is the Oxford Children's Corpus of Writing (OCCW). This corpus is formed of writing produced for BBC Radio 2's *500 Words* short-story writing competition, which is jointly hosted by Oxford University Press. This annual competition asks children aged 4–13 to write fictional, original stories that are no longer than 500 words with the aims of encouraging literacy and creativity across the UK, as well as the development of online learning resources. Since its inception in 2011, the competition has received over a million submissions, and with thousands of submissions every year the corpus is a rich source of

information on children's writing development, their interests, and their new and imaginative use of words (Oxford University Press, 2020). Every year, the entries are gathered, collated, and organised by the Children's Dictionaries & Children's Language department of the Educational Division of Oxford University Press. The team at Oxford Children's Dictionaries uses these corpora to produce annual language reports as well as to choose an annual 'Children's Word of the Year'. These words are chosen based not on overall frequency but on words that have become particularly prominent for various reasons. They may have increased in usage by a noticeable amount or represent a theme that captured the minds of children across the UK. Examples for the past four years are *coronavirus* (2020), *Brexit* (2019), *plastic* (2018), and *refugee* (2017). This vocabulary shows that the stories submitted to 500 Words represent the inventiveness of children and the range of topics that children are engaged by and subsequently want to write about. Competitions such as this are a valuable way for children to develop their out-of-school literacy practices, and can shed light on how children use language to interpret the world around them (Knobel & Lankshear, 2003).

The competition has been running since 2012, and Oxford University Press produces a separate corpus for each year of the competition. Table 5.1 presents the breakdown of the corpora. At the time of designing this study in 2020, the 2019 corpus was the most recent addition to the Oxford Children's Corpus and thus was the corpus used for the more detailed analyses in this chapter. The competition was paused in 2021 and 2022 due to COVID-19. It resumed in 2023, but there was not enough time remaining in the PhD thesis to analyse this dataset.

Table 5.1

Number of Words Per Year in The Oxford Children's Corpus (Writing) (OCCW)

Year	Number of words
2012	32,670,477
2013	39,940,239
2014	50,697,572
2015	50,069,664
2016	54,463,610
2017	54,802,425
2018	55,507,155
2019	46,340,063
2020	47,400,231
Total	**431,891,436**

After 2012 and 2013, the number of words per year is generally stable at around 50 million, which represents over 100,000 different entries annually. In total, over the nine years, there have been 1,040,982 stories submitted to the competition. At the time of writing, the Oxford Children's Corpus of Writing is the largest corpus of children's writing in English in the world, containing approximately 440 million words.

The following metadata labels are available for each entry in the corpus:

- Age (from 4- to 13-years-old);
- Gender (male and female);
- Keystage (KS1, KS2 and KS3);
- Region (119 regions in England, Scotland, Wales and Northern Ireland; by county);
- Wordcount.

Oxford University Press have divided the corpora into sub-corpora based on either age (younger, age 5–9; and older, age 10–13) or gender (male/female). These age bands are the competition age bands. Notably, the competition age bands for the upcoming 2024 competition have been adjusted (to 5–7 and 8–11).

5.2.2 Methodology

The methods employed to interrogate this corpus are as follows:

1. Parts of speech analysis (frequency data);
2. Keyword analysis (using reference corpora);
3. Collocational analysis (using word sketches, concordances, and collocation tools).

These measures all contribute to quantifying the **lexical sophistication** of this corpus to answer the first research question for this study.

5.2.3 Frequency Analysis

The most basic sort of analysis we can do in a corpus is a frequency analysis, where we find its most common words. In this study, frequency information was generated for the entire corpus and then for each part of speech separately, to produce the more fine-grained analysis as recommended by previous work (Durrant et al., 2021). These frequencies are generated as occurrences per million words and are presented as percentages of the total corpus in the results sections. The Oxford Children's Corpus is tagged for parts of speech using the Oxford English Corpus tagset, which is described on the Sketch Engine website [https://www.sketchengine.eu/oxford-english-corpus-tagset/ accessed 24/04/2023]. When producing this frequency information, Sketch Engine uses lemmatised word lists, and thus when calculating percentages the parts of speech were divided by the total number of lemmas in each corpus. The search parameters for the word list function were: case insensitive, minimum frequency = 5, exclude non-words (tokens which do not start with a letter).

In linguistics, there is an ongoing discussion on which lexical unit is most appropriate for research and pedagogy (D. Brown et al., 2021; Webb, 2021). The choice of whether to define words at the lemma or word family level is a key methodological consideration in lexical analyses. Theoretically, the question is whether having knowledge of one item can predict knowledge of the

other items in the group; practically, grouping more tokens together will affect calculations, i.e., lower scores in measures of lexical diversity. The lemma level seems to be the most appropriate as there is evidence that learners who can recognise one member of a lemma are usually able to recognise other members, but that this is not the same for word families (D. Brown et al., 2020); using lemmas may also allow for greater precision and fewer assumptions about knowledge, especially in corpus linguistics where the tools used can easily find and count lemmas (Gablasova & Brezina, 2021). Webb (2021) argues that there is no 'lemma dilemma' due to the fact that the majority of software tools are lemmatised and using lemmas is more precise and requires fewer assumptions about the morphological and semantic knowledge of learners. For this research, lemmas were chosen as the most appropriate unit as we can assume that these children have some working knowledge of the verb and noun inflection system (Kyle, 2019).

5.2.4 Keyword Analysis and Reference Corpora

Keyword analysis is used to provide information about a corpus. Essentially, a keyword is a word that occurs more frequently (statistically speaking) in one corpus when compared to another. A keyword analysis using Sketch Engine has been described as a corpus-driven approach, as it involves finding out statistically more frequent words in one corpus compared to another; the researcher does not form prior hypotheses about the language used. Keywords often have to be investigated in context, by also looking at concordance and collocation information. The keyword function of Sketch Engine compares a focus corpus and a reference corpus using statistical formulas which are described in detail on the Sketch Engine's website; for further information on both keywords and the Sketch Engine in general, see Kilgarriff (2009), Wild et al. (2013), and Kilgarriff et al. (2014).

Using reference corpora is a common way to investigate differences between two or more corpora and allows the investigation of keywords, providing important information for measuring

lexical sophistication (Durrant, 2022a). Generally, reference corpora are split into two types: *input corpora* and *target corpora*. An input corpus represents what a group has likely been exposed to, and a target corpus exemplifies the sorts of language a group is 'aspiring' to. For example, for analysing L1 writing development in children, an input corpus could be a collection of children's books (content that children are likely to consume) and a target corpus could be academic texts (assuming that education in the UK is preparing children to produce these sorts of texts). In this study, three corpora are used for making comparisons. Their size, genre and audience are summarised in Table 5.2.

Table 5.2

Reference Corpora

	Size (lemmas)	**Genre**	**Audience**	**Corpora type**
BNC (Written imaginative)	16,341,915	Fiction	Adult	Target
Oxford English Corpus	2,087,793,021	General	Adult	Target
Oxford Children's Corpus Reading (2017)	63,289,908	Fiction and non-fiction	Children	Input

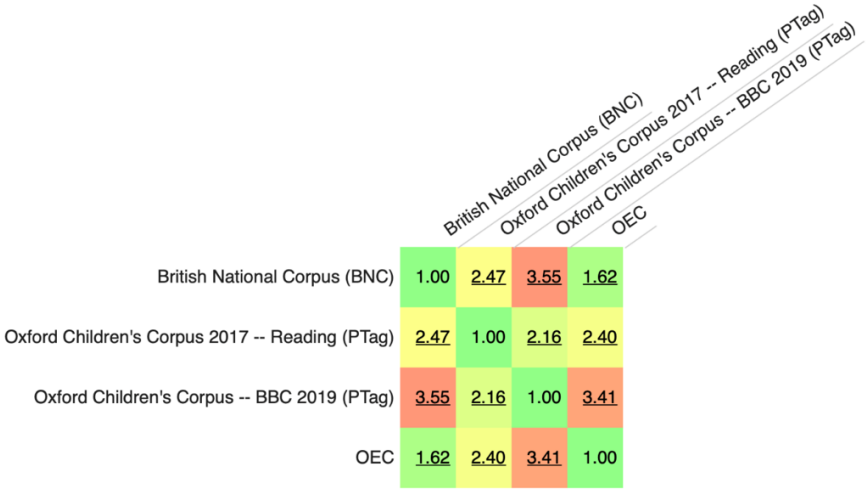
Large national corpora are often used as a proxy input due to difficulties in sourcing large, representative corpora of child learner input, but using such generalised corpora can lead to problems when calculating the relevant distributions of vocabulary items because real learner input is narrower than a national corpus (i.e., a generalised input corpus may not actually represent the items that learners are being exposed to such as social media, cinema, television, etc). In this study, the Oxford Children's Corpus of Reading (OCCR; described above) is used as an input corpus. The most recent corpus available at the time of carrying out this analysis (2019) was used in order to present the most recent available data. Using the OCCR allows us to compare writing *for* children to writing *by* children and gives us an idea of how children are being influenced by texts they are likely to be consuming.

Choosing the target corpus is more difficult. Durrant (2022) describes two types of target corpora: a peer target corpus and a prospective target corpus. The former represents writing deemed as successful produced by learners at the same level as the group being investigated. In Study 2 (Section 4.3), a bespoke corpus was created using the stories produced by the winners of the 500 Words competition. The second type of target corpus is a prospective target corpus and describes corpora containing texts which learners "might realistically hope to produce at some point in the future" (Durrant, 2022a, p. 50). A prospective target corpus allows us to trace development across ages by having a fixed point of reference. Choosing a prospective target corpus involves identifying what sorts of writing we are hoping that the children who have submitted stories to this competition might produce in the future. For this case, the written imaginative domain sub-corpus of the British National Corpus (BNC) was chosen. This is a corpus of English fiction (novels and short stories) written by and for adults during the latter half of the 20th century. Using the BNC gives us a prospective target, i.e., in that children producing these short stories may eventually be aiming to produce written fiction of this standard. It allows us to examine the differences between writing by/for adults and writing by/for children. The second comparative corpus is the Oxford English Corpus (OEC), which is a generalised English written corpus that takes its data from the web. This acts as a sort of "control" corpus, allowing comparison of the other corpora to generalised English.

Sketch Engine provides a corpus comparison tool to demonstrate how similar two or more corpora are. Figure 5.1 demonstrates how the reference corpora used in this study differ from each other.

Figure 5.1

Comparison Of The Reference Corpora Used In This Study



The greener the square, the more similar the corpora are, and the redder the square, the more dissimilar. Thus we can see that the OCCW is most similar to the OCCR, as expected, and most dissimilar from the BNC followed by the OEC. Sketch Engine calculates these figures as follows (see https://www.sketchengine.eu/guide/compare-corpora/ accessed 1 June 2024):

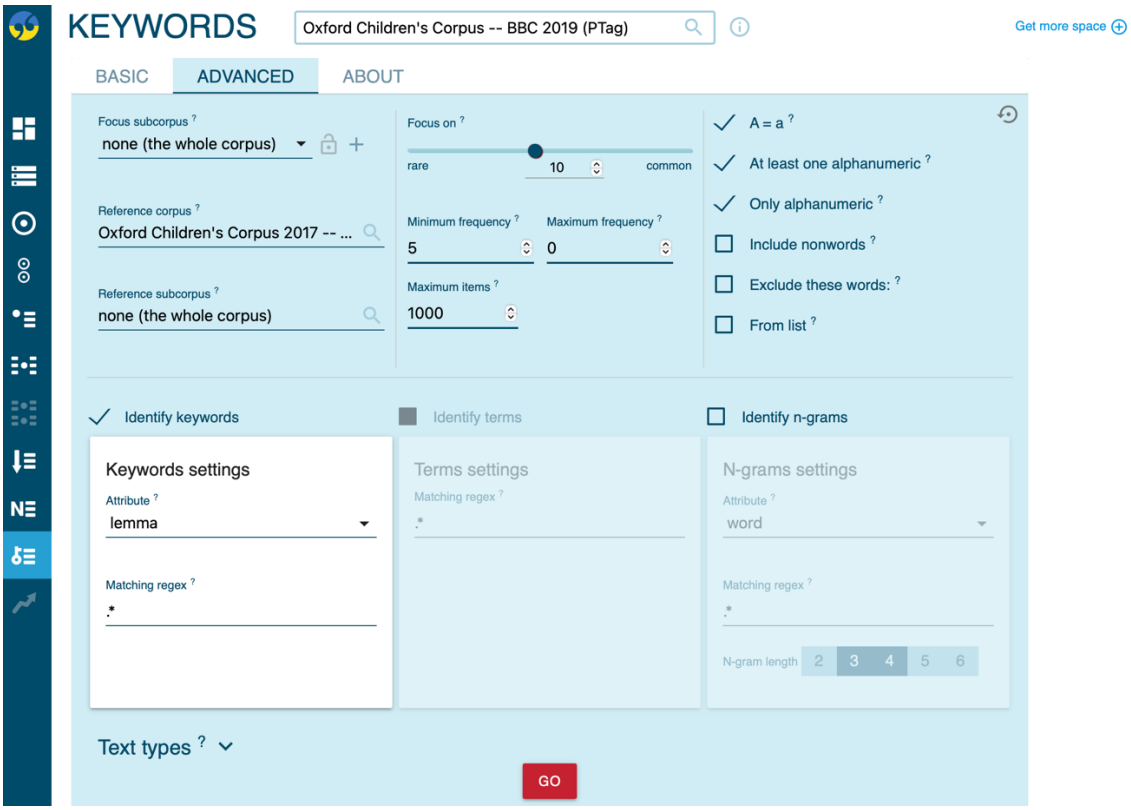
1. The keyness score for every word is computed. The corpus with a higher relative frequency of the word is set as the focus corpus. The corpus with a smaller relative frequency of the word is used as the reference corpus. Thus, the resulting number is always more than 1 or 1 in case the frequency is the same.
2. The 500 words with the highest keyness score are identified.
3. The arithmetic mean (average) is calculated from the keyness scores of the top 500 words. The result expresses the similarity of the corpora. This is the number displayed in the chart on the corpus comparison result screen.

The keyword comparison tool is a key feature of Sketch Engine. Figure 5.2 depicts the keywords analysis screen in Sketch Engine. The focus corpus, i.e., the OCCW 2019 corpus, is at the

top. The reference corpus in this example is the Oxford Children's Corpus (Reading) 2017 Ptag (tagged for parts of speech). Sketch Engine allows adjustment to focus on either higher or lower frequency words using their add-N parameter (Kilgarriff, 2009). For the keyword analyses in this study, this parameter was set at 10, giving a very slight preference for rarer words. This gave the best balance between rare and common words. Setting the slider all the way to the left (rare) meant that the results were generally things like names or typos, whereas moving it to the right meant that the results generated highly frequent words in both the focus and reference corpora, such as *be, they, she, go, my*, etc., and thus were uninteresting. The results are organised by their keyness score computed using the simple maths method, which compares the likelihood of encountering a word in one corpus with another.

Figure 5.2

Keyword Analysis in Sketch Engine



5.2.5 Collocational Analysis

One of the main features of Sketch Engine is its word sketch tool, which allows a user to search for a particular word to view a one-page summary of its grammatical and collocational behaviour. In this way one can see how a word is used in context, across different genres, and can compare a word across different corpora, if desired. Word sketches and the concordance analysis in Sketch Engine organise their collocations by a score calculated by the logDice algorithm, which expresses the typicality (or strength) of the association while remaining unaffected by the size of the corpus (Rychlý, 2008). A typicality score is generated by ranking more 'flexible' associations as weaker than more specific associations. For example, when looking at collocates for *bedroom*, a *small bedroom* would be a weaker collocation than a *twin bedroom*, because the adjective *small* is very frequent and can be combined with a large number of other nouns. An example of the word sketch screen is seen in Figure 5.3 for the word *dog* in the OCCW 2019 corpus. A number of useful pieces of information are included in this word sketch. We can see that the frequent collocations of *dog*, such as *cats and dogs*, *pet dog*, or *puppy dog eyes*; the word sketch also breaks these

collocations down for different parts of speech, so that we can see what verbs or prepositions often modify *dog*, for example.

Figure 5.3

Word Sketch Example For 'dog' in the Oxford Children's Corpus Of Writing (2019)

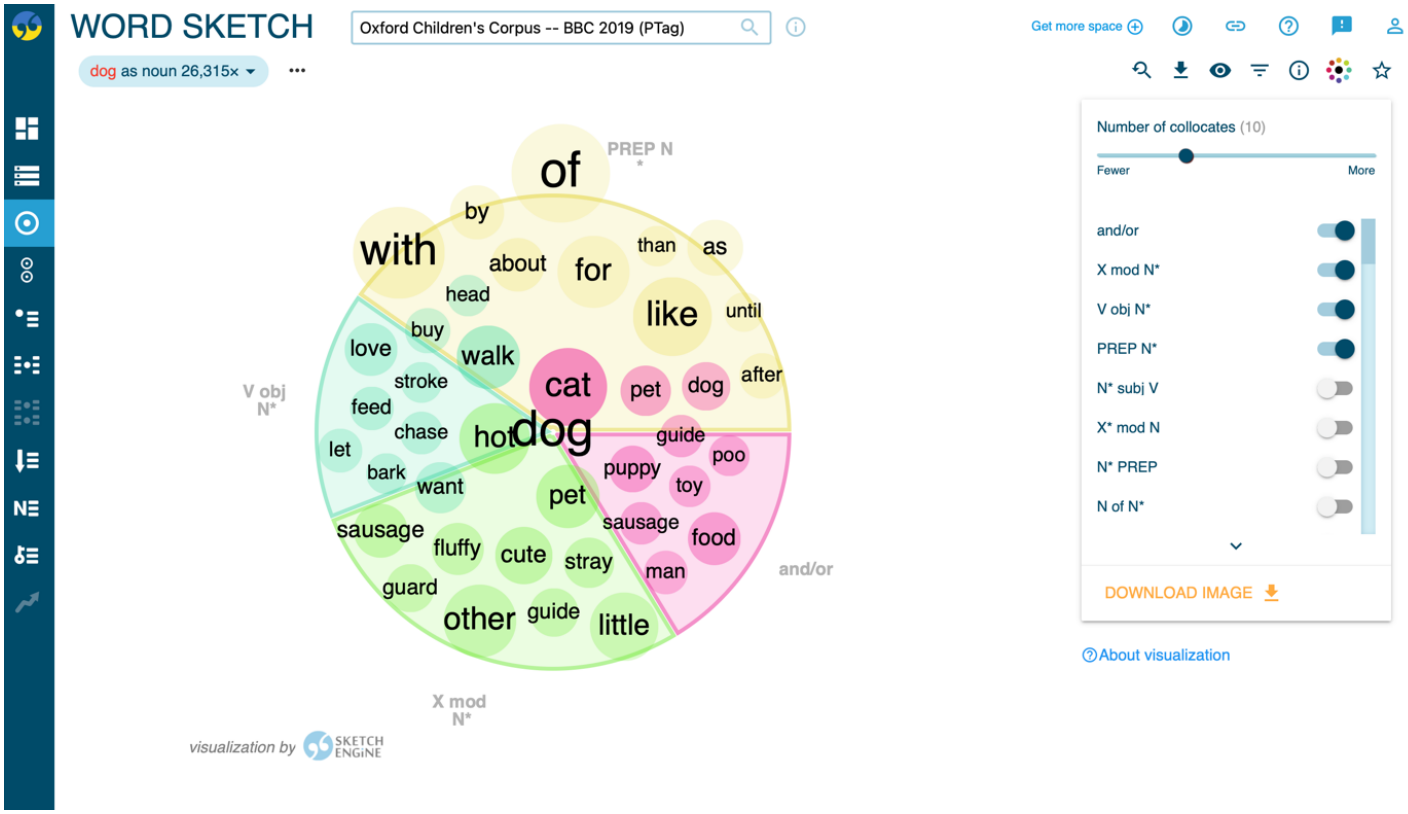
Sketch Engine's word sketch features also enables data visualisation, where more common words generate bigger circles. An example for *dog* is presented in Figure 5.4. There are four key features to be aware of when studying these visualisations: (1) the distance from the centre indicates the typicality score, e.g., *hot dog* is more typical than *other dog*; (2) the size of the circle indicates frequency; (3) the circle colour indicates which grammatical category the collocation belongs to; (4) the segment size indicates the size of the grammatical relationship in relation to the other visualised relationships, i.e., how many collocations it contains in total. In this example, there are more instances of prepositions modifying *dog* (noun) than cases where a verb is paired with the

noun *dog* as an object. These visualisations are used in this study, modelling Wild et al. (2013), to exemplify how some particular words are used in children's writing.

It is important to note that there are several parameters which can be adjusted when producing these visualisations, such as the number of collocates produced and how they are presented. In this research, Sketch Engine's default parameters were used.

Figure 5.4

Word Sketch Visualisation Example For 'dog' In The Oxford Children's Corpus Of Writing



Finally, Sketch Engine allows one to search for concordances of a particular word. When these are presented in this study, a random sample of lines was taken (set at Sketch Engine's default of 200) using Sketch Engine's 'get a random sample' feature.

5.2.6 Results

This section will present the results of the investigation into the Oxford Children's Corpus. The most recent corpus available at the time of writing is the 2020 BBC 500 Words Corpus. It is pertinent to check whether this corpus is representative of the rest of the corpora. Figure 5.5 displays how similar each year's corpus is to the others, and shows that they are all fairly similar in regard to keywords. A score of 1 indicates that the corpora are identical. This figure shows that the OCCW corpora are all fairly similar, with the largest difference seen between the oldest and most recent corpora. If there had been large deviations, further investigation would have had to occur to investigate why one particular corpus was less similar than others.

Figure 5.5

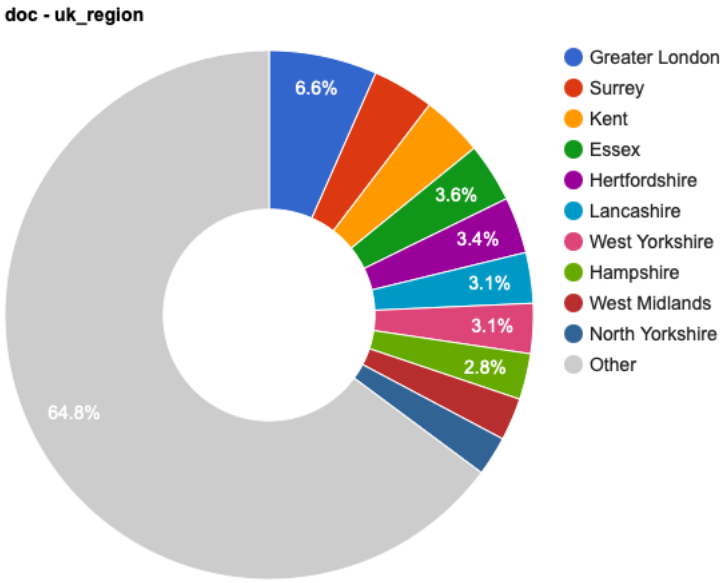
Screenshot Showing Results From The 'Compare Corpora' Tool In Sketch Engine, Attribute: 'Lemma'. Only Six Of The OCCW Corpora Were Available For This Comparison

	Oxford Children's Corpus -- BBC 2016 (Ptag)	Oxford Children's Corpus -- BBC 2017 (Ptag)	Oxford Children's Corpus -- BBC 2018 (PTag)	Oxford Children's Corpus -- BBC 2019 (PTag)	Oxford Children's Corpus -- BBC 2020 (PTag)
Oxford Children's Corpus -- BBC 2016 (Ptag)	1.00	1.09	1.09	1.10	1.12
Oxford Children's Corpus -- BBC 2017 (Ptag)	1.09	1.00	1.09	1.09	1.11
Oxford Children's Corpus -- BBC 2018 (PTag)	1.09	1.09	1.00	1.09	1.10
Oxford Children's Corpus -- BBC 2019 (PTag)	1.10	1.09	1.09	1.00	1.09
Oxford Children's Corpus -- BBC 2020 (PTag)	1.12	1.11	1.10	1.09	1.00

Next, we can present the demographic data of this corpus. If we wish to investigate the word gap, and the idea of the word gap is predicated on the influence of poverty on the language environment of children, it is interesting to examine the region distribution within the Oxford Children's Corpus of Writing. Each text in this corpus is tagged for 'uk_region', which separates the texts into UK counties. The following figure is the output from Sketch Engine when you ask it to show you the distribution of regions within the corpus. The 'Other' section in this figure includes all other counties in the UK grouped together, as they each represented <1% of the total corpus.

Figure 5.6

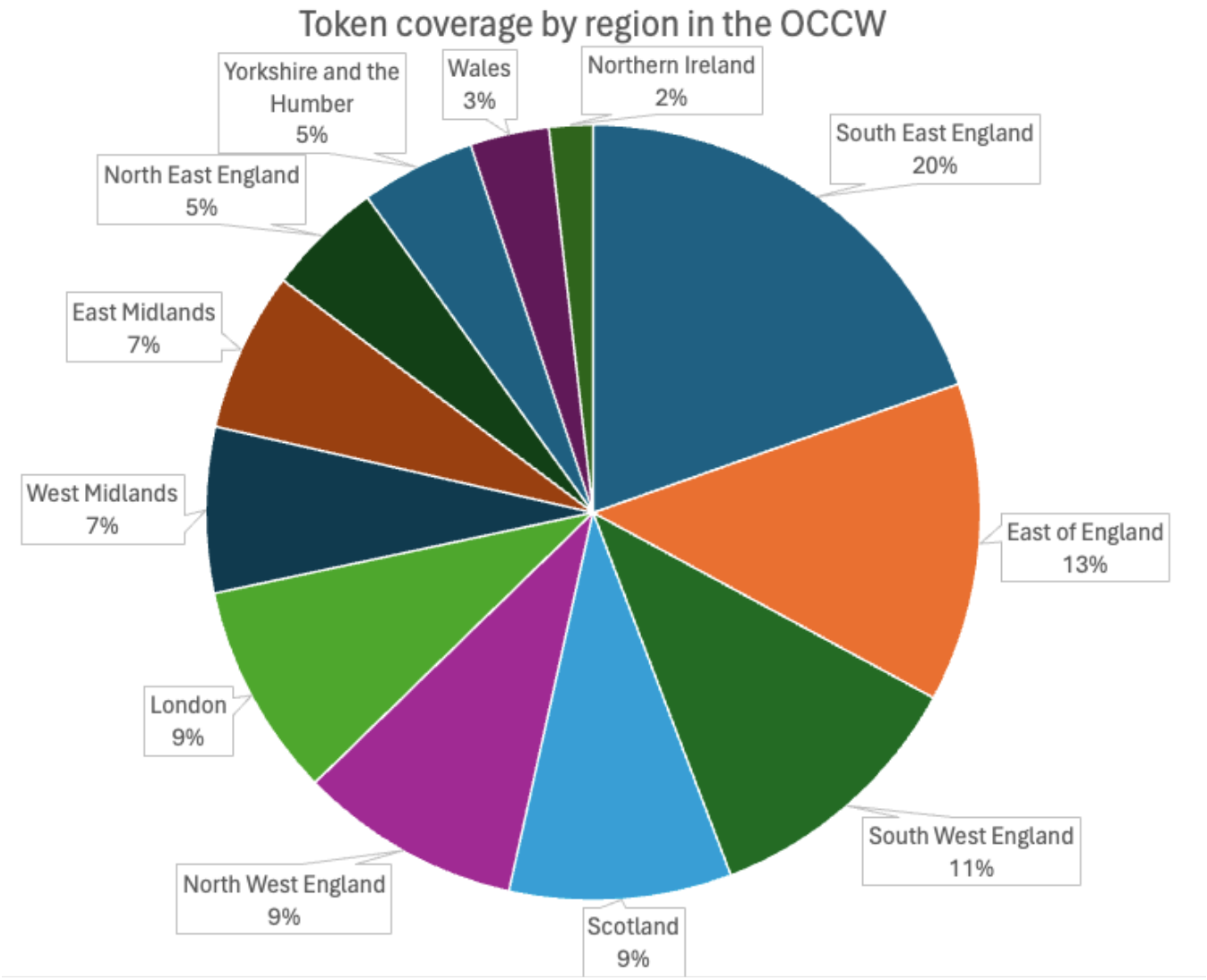
Region Distribution as Tagged In The OCCW



It is useful to be able to separate the data in this way if one were interested in evaluating the vocabulary used in a particular county, for example. However, statistics on poverty in the United Kingdom are often based on region. For example, see the latest poverty statistics on the UK by the Joseph Rowntree Foundation (https://www.jrf.org.uk/uk-poverty-2024-the-essential-guide-to-understanding-poverty-in-the-uk accessed 6 Jun 2024). Thus, these numbers from Sketch Engine were downloaded into a .csv file and imported into Excel. Using a list of counties and their regions obtained from github (https://gist.github.com/radiac/d91d2ed1b971c03d49e9b7bd85e23f1c accessed 6 Jun 2024), the counties drawn from Sketch Engine could be grouped into regions.

Figure 5.7

Token Coverage By Region In The OCCW (2020) Corpus



Most entries in the corpus are from South East England, followed by the East of England and South West England. The following table presents poverty rate figures in percentages, as determined by the Joseph Rowntree Foundation, using figures from the Department of Work and Pensions, which operationalises poverty as number of households below average income.

Table 5.3

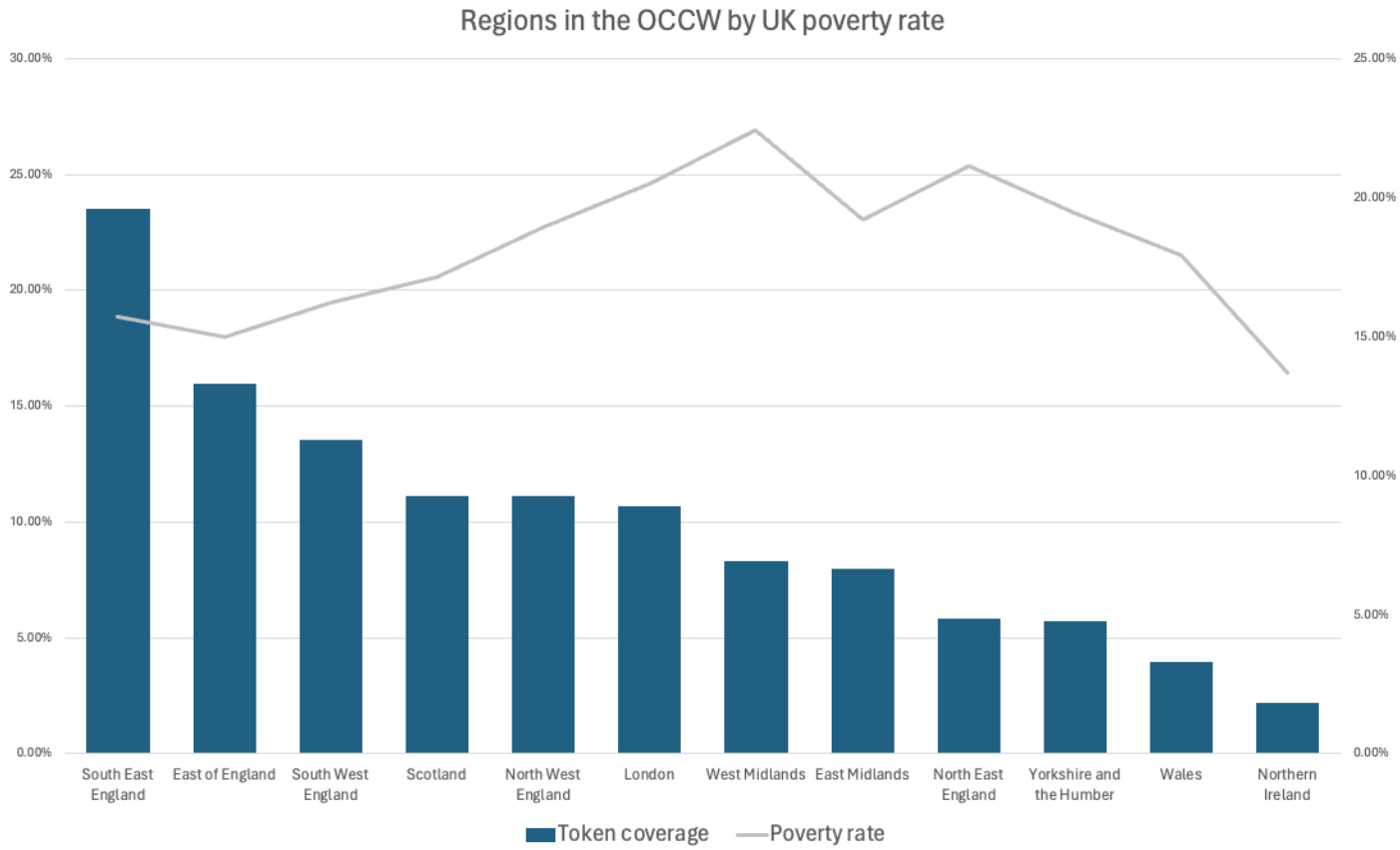
Poverty Rates in the UK

Region	**Poverty rate**
West Midlands	26.9%
North East England	25.4%
London	24.6%
Yorkshire and the Humber	23.3%
East Midlands	23.0%
North West England	22.7%
Wales	21.5%
Scotland	20.5%
South West England	19.5%
South East England	18.9%
East England	18.0%
Northern Ireland	16.5%

Putting this information together, we can see that the general trend is that the texts in the OCCW tend to come from area in the UK with lower poverty rates (Figure 5.6). There are fewer submissions from the areas with higher poverty rates, such as the West Midlands and London.

Figure 5.8

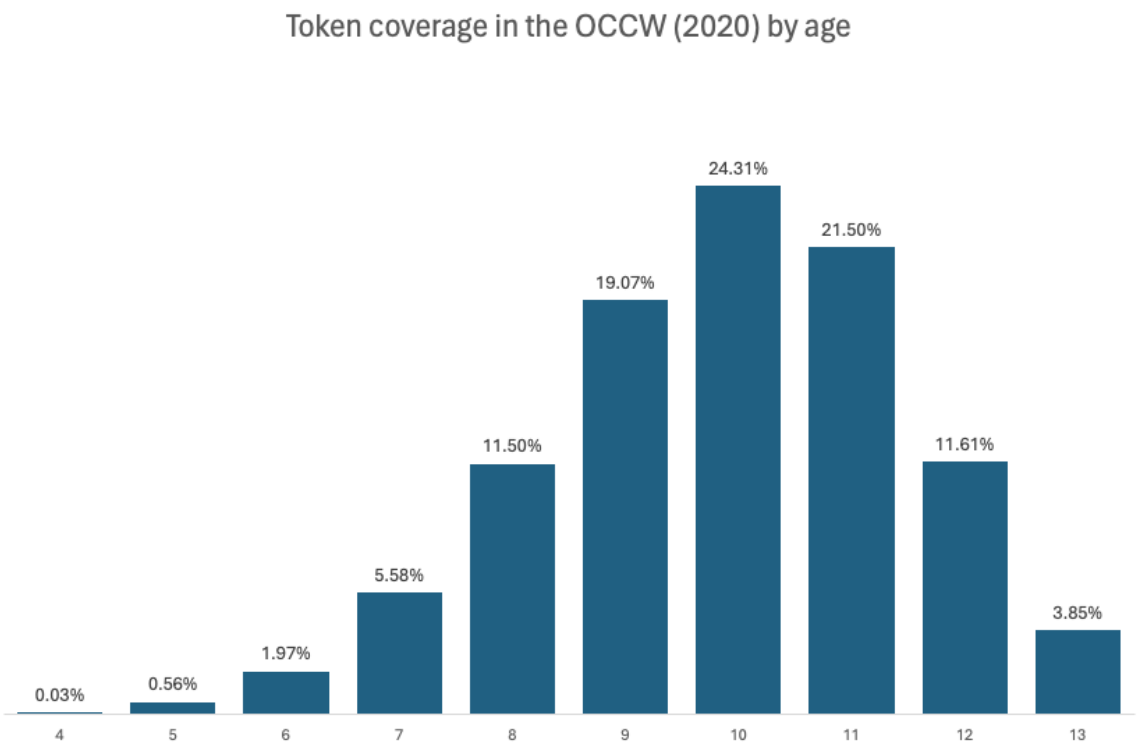
Regional Distribution In The OCCW By UK Poverty Rate



In terms of gender, 58.7% of the texts in the OCCW (2020) were tagged as female and 41.3% as male. In terms of age, in the OCCW (2020) corpus the majority of texts are written by children aged 9, 10, and 11. The BBC 500 Words competition has changed the way it organises the competition by age over the years. At the time of writing, the age bands were 4–9 and 10–13. However, as we can see by Figure 5.7, very few entries come from children aged 4, 5, 6, or 13. In the next iteration of the competition, which commences in September 2024, the age categories are 5–7 and 8–11.

Figure 5.9

Token Coverage in the OCCW by Age



5.2.6.1 Nouns

The twenty most common nouns for all four corpora are presented in Table 5.4. The most common nouns in the OCCW can be sorted into several thematic groups: people (*friend, man, people, girl, boy*), parts of the body (*eye, head*) nature (*tree, life*), time expressions (*day, time, year, one*), places (*door, house, school, thing, room*), and other (*thing, way, name*). The nouns that occur in the most common 20 in the OCCW but not in the OCCR are *friend, school, room, tree, one, life*, and *name*. In fact, *friend, school, tree, one* and *name* are nouns unique to the most common twenty nouns for the OCCW, indicating that these are words that children particularly like to use in their writing despite them not being as common in their reading materials.

Furthermore, noticeable changes occur in the adult-directed corpora. The top twenty nouns in the BNC include *woman*, *nothing*, *night*, and *voice* appear as distinguishing words in adult fiction.

The top eight nouns in the OEC do appear in the OCCW, but the rest of the nouns do not. This may be due to a focus on non-fictional adult topics such as business, news, and politics (e.g. *government, company, country, case*). The OEC includes no body parts, which are a common semantic group in the other corpora. The OCCR and OEC both include the word *child*, but this word is not common in the OCCW, suggesting that even though children write stories where children are often the protagonist, they do not explicitly refer to either *child* or *children*. In the table below, asterisks indicate that the word was unique for being in the top 20 for that corpus.

Table 5.4

Top 20 Most Common Nouns In The OCCW (2020), OCCR (2019), BNC, And OEC

	OCCW		OCCR		BNC		OEC	
#	Item	%	Item	%	Item	%	Item	%
1	day	0.21%	time	0.20%	time	0.21%	people	0.18%
2	time	0.21%	man	0.17%	man	0.20%	time	0.17%
3	friend*	0.16%	way	0.12%	eye	0.14%	year	0.16%
4	door	0.16%	day	0.12%	hand	0.14%	way	0.10%
5	house	0.15%	hand	0.12%	way	0.14%	day	0.09%
6	eye	0.14%	thing	0.11%	thing	0.12%	thing	0.08%
7	school*	0.13%	eye	0.11%	day	0.12%	man	0.08%
8	man	0.12%	people	0.10%	face	0.11%	life	0.07%
9	people	0.12%	head	0.10%	head	0.10%	work*	0.07%
10	thing	0.12%	year	0.08%	door	0.10%	part	0.07%
11	room	0.12%	face	0.08%	something	0.10%	world	0.07%
12	girl	0.12%	child	0.08%	room*	0.09%	child	0.06%
13	way	0.12%	house	0.07%	woman	0.09%	country	0.06%
14	year	0.11%	boy	0.07%	house	0.08%	number	0.06%
15	boy	0.10%	door	0.07%	people	0.08%	government*	0.06%
16	tree*	0.09%	something	0.07%	nothing	0.08%	company*	0.06%
17	one*	0.09%	place	0.07%	year	0.08%	case*	0.06%
18	life	0.09%	water*	0.07%	life	0.07%	group*	0.06%
19	name*	0.09%	world	0.07%	night	0.07%	place*	0.06%
20	head	0.09%	girl	0.07%	voice*	0.07%	woman	0.06%

Friend is the third most common noun in the OCCW corpus and is the first noun which is *not* also in the top twenty for the other corpora. For this reason, it was chosen for further analysis. Figure 5.8 presents its word sketch and some concordance line examples. Its most frequent collocate

was *good*, used the majority of the time in the phrase *best friend*, followed by *new* and *my*. The most frequent verbs found near friend are *have*, *make*, and *become*, and friends often *go*, *come*, *call*, and *say*. *Friend* is often pluralised and often occurs with possessive structures. These uses of the word *friend* emphasise how children experience their friendships: in an active and positive manner, marked by direct involvement in the world around them in which they move.

Figure 5.10

Word Sketch Visualisation and Concordance Lines For Friend

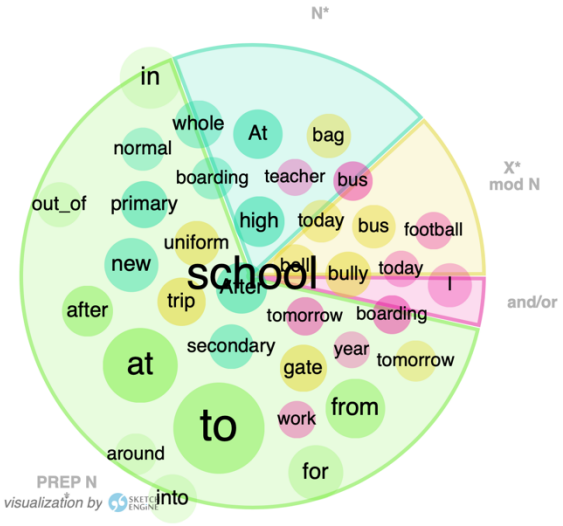
><s> Intigo went in his spaceship to bring his alien	**friends**	to rule the empty Earth. </s><s> But on his way
mess, let me help you". </s><s> Together the new	**friends**	began to clean the kitchen and make a new batc
afternoon they were at the park and they saw their	**friends**	, Ted the tiger and Leo the Lion and they were o
show she had told some friends, so they told some	**friends**	and so on but now she was feeling rather nervou
:s> He lived in a big mansion all by himself, had no	**friends**	and lived in the middle of nowhere. </s><s> It wa
e all counting on him to save them all. </s><s> His	**friends**	were cheering his name "Michael" "Michael" "Mi
s in a year. </s><s> I went to sit on a table with my	**friends**	and placed my bowl in front of me. </s><s> My b
Facebook when suddenly I got a text from my best	**friend**	saying" hey Lucy did you know that Samantha is
/s><s> I went back up onto my boat and asked my	**friend**	, "Did you bring all of the urns to drop in?" </s><

School is another noun that is especially relevant to children, who spend a large proportion of their time in school, and was unique to being in the most frequent nouns in this corpus. Despite this, *school* is not one of the most common words used in texts directed at children, i.e., reading materials present in the Oxford Children's Corpus of Reading. This suggests that children may enjoy

writing about school because it directly relates to them and their life experiences, even if they do not necessarily often read about the same experiences. Figure 5.9 presents the word sketch for *school*, followed by some example concordance lines. From this we can see that their writing exemplifies different types of school (*high school, boarding school, primary school, secondary school or new school*), and that children often write about going *to* or *from* school. Ideas related to school are also present, such as school *bus*, school *trip*, or school *run*. Again, this noun represents something important to children which they then want to write about.

Figure 5.11

Word Sketch Visualisation and Concordance Lines For School



ive. </s><s> Cat said the same thing every time Anne was off to **school** , "Go and smash 'em"! </s><s> Anne could not help feeling like
t wait to tell Cat, Connor and Jack! </s><s> Jack was not on the **school** bus the following day and for a whole week after. </s><s> He se
/hole week after. </s><s> He seemed to have disappeared from **school** too, but on the day of the hockey match with St Catherine's Prep
oo, but on the day of the hockey match with St Catherine's Prep **School** , Jack was cheering from the side line but Cat and Connor's che
do again and again. </s><s> She was girl of the match and her **school** won the fixture. </s><s> Jack disappeared before Anne could sa
pharmacist delivers like post man pet </s><s> Nanas not on the **school** run but she does the hospital run and comes home to her cosy t
ar no smoke in sight. </s><s> All of the village juniors went on a **school** trip to the nether. </s><s> Mr pigclawmen (who is a horrid pigme
ca dressed Mr Fart Fingers up as a teacher and took him to her **school** . </s><s> Rebecca said </s><s> "My teacher is really rude and i

5.2.6.2 Verbs

The top twenty most common verbs in all four corpora are presented in Table 5.4. This table shows us that the OCCW presents a very similar profile to the other corpora, with the exception of *run, start* and *call*, which are only most common in the OCCW. The most common verbs in the OCCW can be grouped into semantic groups that are indicative of narrative texts, such as moving and existing (*be, go, come*), performing actions (*say, do see, look, find, run, call, tell, make*), and verbs that involve mental actions (*think, know*). Many of the verbs common in children's writing are similar to the verbs used in other sources of writing (e.g., *say, exclaim*), indicating a "pool" of verbs

common amongst different types of written English. In general, the common verbs in the OCCW represent the 'core' processes in narrative texts, and shows children's preoccupation with movement, physical space, and direct speech (Thompson & Sealey, 2007; Wild et al., 2013). The BNC, on the other hand, includes verbs that do not occur in other corpora (*want, feel, ask and leave*), and which suggest a focus in adult fiction on emotional rather than physical states. In the table below, asterisks indicate that the word was unique for being in the top 20 for that corpus.

Table 5.4

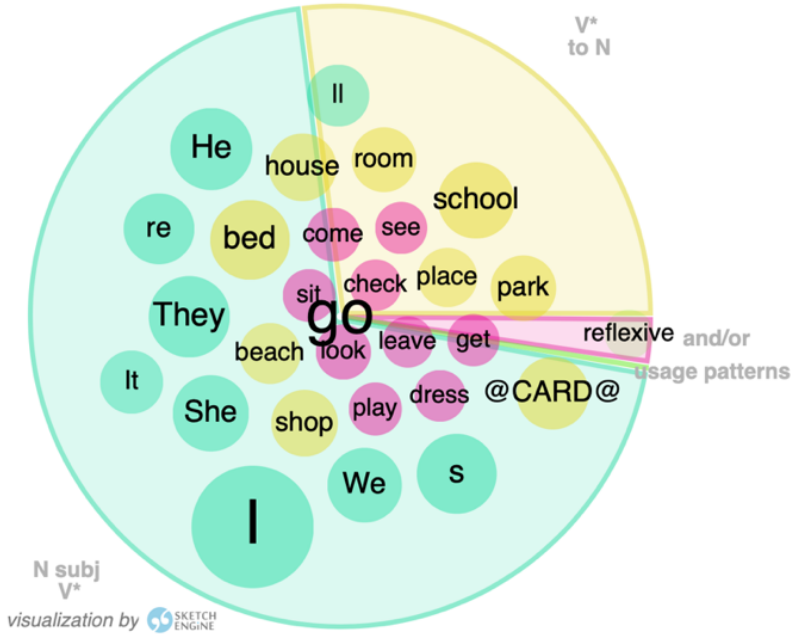
Top 20 Most Common Verbs In The OCCW, OCCR, BNC (Written Imaginative), And OEC

	OCCW		OCCR		BNC		OEC	
	Item	%	Item	%	Item	%	Item	%
1	be	4.12%	be	3.61%	be	4.18%	be	4.03%
2	have	1.14%	have	1.28%	have	1.76%	have	1.24%
3	go	0.81%	do	0.60%	do	0.87%	do	0.49%
4	get	0.61%	say	0.57%	say	0.67%	say	0.39%
5	say	0.60%	go	0.36%	go	0.40%	will	0.35%
6	do	0.50%	see	0.29%	know	0.34%	would	0.25%
7	see	0.45%	will	0.28%	see	0.30%	can	0.23%
8	come	0.33%	would	0.28%	think	0.29%	make	0.22%
9	look	0.30%	make	0.26%	get	0.29%	go	0.21%
10	could	0.29%	get	0.25%	look	0.27%	get	0.20%
11	know	0.28%	come	0.25%	come	0.26%	take	0.17%
12	find	0.25%	know	0.24%	make	0.23%	see	0.15%
13	think	0.24%	could	0.23%	take	0.22%	know	0.15%
14	would	0.23%	can	0.22%	tell	0.18%	could	0.13%
15	run*	0.23%	look	0.21%	want*	0.17%	come	0.13%
16	make	0.23%	take	0.20%	feel*	0.15%	think	0.13%
17	start*	0.21%	think	0.20%	ask*	0.13%	use*	0.13%
18	take	0.20%	find	0.14%	give	0.13%	give	0.11%
19	call*	0.18%	tell	0.13%	find	0.13%	find	0.10%
20	tell	0.16%	give	0.13%	leave*	0.12%	should*	0.10%

Figure 5.10 presents the word sketch and some concordances of *go*, and shows the main uses: N + *went*, *going to* + V, *go to school*, and *go to sleep*. Children often write about *going to bed*, *going home*, or *going to school*, *the beach*, *shops*, or *parks*.

Figure 5.12

Word Sketch Visualisation and Concordance Lines For Go



› together to turkey we had so much fun just being a family, we	went	swimming and on boat trips and went out to eat we had so muc
e middle of the night, there was a family from the city who was	going	to move to the countryside because they were a little poor, that
and hasn't treated people she should. </s><s> Tomorrow I am	going	to teach her how to be nice. </s><s> I can't even get to where i
<s> I said yes because I love plaing whith them. </s><s> "Now	go	!" </s><s> I heard a rather high pitched voice shout from inside
t was gone. </s><s> I breathed a sigh of relief. </s><s> Before	going	to sleep, I threw the book into the fire. </s><s> But what I didn't
/s><s> "I hope it does." </s><s> Replied Collin. </s><s> Alfred	went	home that night and thought about Collins dream. </s><s> He s
</s><s> It got to 21:30 and he turned off his PS4. </s><s> He	went	down stairs to get some supper and he had two delicious piece
> "I don't actually know?" </s><s> replied Pippa. </s><s> They	went	to go and check out what the other pigs were doing, and they s
f she did do what she was going to do she would be looking to	go	to a different school. </s><s> Four hours later and the school d

Get is another common verb in the OCCW that is more common than in the other corpora. A word sketch and concordance investigation show that get is commonly used reflexively, with other verbs such as *try and get*, *go and get*, or with nouns such as *get revenge*, *get food*, or *get ideas.*

5.2.6.3 Adjectives

The adjectives used in the OCCW overlap less with the other corpora. Although the top ten adjectives are similar, individual differences are apparent. For example, *big* is used twice as much in

the OCCW (0.11%) compared to the OCCR (0.05%), and *dark* is present in the top ten adjectives in the OCCW but only appears far down the list in the BNC, and not at all in the OCCR or OEC. After the top 10, further differences between the OCCW and the other corpora emerge. *Black, bad, happy, red, blue*, and *ready* are found in the OCCW's top 20 adjectives but are not this frequent in the other corpora. In fact, colours do not appear in any other corpora. Upon further investigation into the OCCR, *white* is the first colour at #25. *Great, own, many, young, such, sure, large, more*, and *high* are present in the OCCR but not the OCCW. In the table below, asterisks indicate that the word was unique for being in the top 20 for that corpus.

Table 5.5

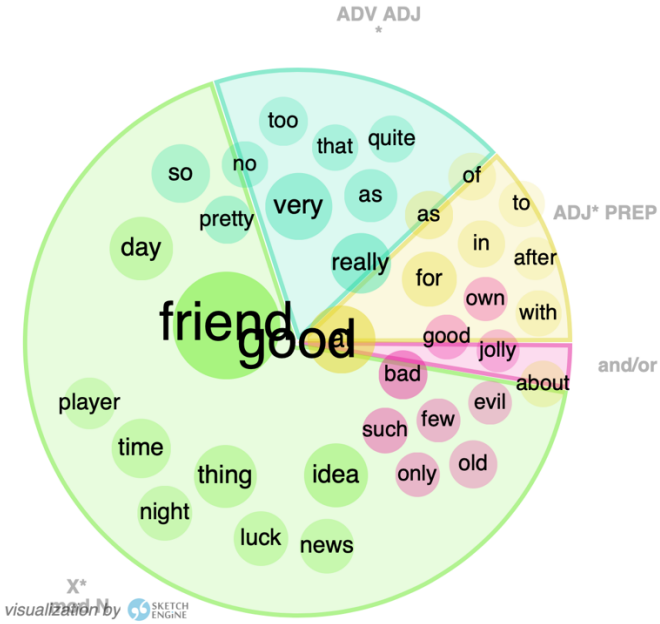
Top 20 Most Common Adjectives In The OCCW (2020), OCCR (2019), BNC (Written Imaginative), And OEC

#	OCCW		OCCR		BNC		OEC	
	Item	%	Item	%	Item	%	Item	%
1	good	0.12%	little	0.13%	good	0.15%	other	0.14%
2	old	0.11%	good	0.13%	other	0.12%	good	0.13%
3	big	0.11%	other	0.11%	little	0.10%	new	0.12%
4	little	0.09%	old	0.10%	old	0.09%	first	0.10%
5	other	0.08%	great	0.09%	more	0.09%	many	0.08%
6	new	0.07%	first	0.08%	own	0.08%	such	0.07%
7	dark	0.06%	long	0.07%	first	0.07%	great	0.06%
8	first	0.06%	new	0.06%	last	0.07%	same	0.06%
9	long	0.06%	own	0.06%	long	0.07%	high	0.06%
10	small	0.06%	many	0.06%	small	0.06%	own	0.05%
11	black*	0.06%	same	0.06%	few*	0.06%	more	0.05%
12	bad*	0.05%	young	0.05%	young	0.06%	big	0.05%
13	happy*	0.05%	small	0.05%	same	0.05%	different*	0.04%
14	same	0.05%	last	0.05%	great	0.05%	old	0.04%
15	red*	0.05%	such	0.05%	next	0.04%	large	0.04%
16	blue*	0.04%	big	0.05%	new	0.04%	small	0.04%
17	ready*	0.04%	sure	0.04%	sure	0.04%	young	0.04%
18	next	0.04%	large	0.04%	many	0.04%	local*	0.04%
19	scared*	0.04%	more	0.04%	much*	0.04%	important*	0.04%
20	last	0.04%	high	0.04%	dark	0.04%	second*	0.04%

Figure 5.9 depicts the word sketch and concordance of *good*, the most common adjective in the OCCW. Children use this word mainly in combination with friend, as in *best friend*, as noted above, but they also like writing about *good luck*, *good ideas*, or *good players*. Things are described as *really good*, *very good*, or *pretty good*, as well as *good and bad* or *good and evil*.

Figure 5.13

Word Sketch Visualisation and Concordance of Good



blue the tree house door appears billy says his	good	byes, turns the handle anti-clockwise and goe
M.E. </s><s> I sighed. </s><s> This was all too	good	to be true. </s><s> There was no universe tha
ad a bump on the head and sweet ten is always	good	for that when George selected to shrink small
oe my mouth, because it feels right. </s><s> It's	good	. </s><s> So I do it again. </s><s> Then I real
gh peoples legs. </s><s> There would be top 50	best	swimmers for the first semi-final competition.
rls woke up to the smell of breakfast(it wasn't a	good	smell) so they ran down the creaky stairs and

5.2.6.4 Adverbs

The most common adverbs present some overlap between the OCCW and the other corpora, but the percentage values show some key differences. *Up*, the most common adverb in the OCCW, represents 0.43% of the corpus, and while it is the most common adverb in the OCCR, it only

represents 0.31% of that corpus; it is not present in the top 20 adverbs in the BNC but is the second most common adverb in the OEC. *Off*, *home*, *really*, and *still* occur in the OCCW but not in the OCCR. *Too* is a common adverb in the three corpora except the OCCW, suggesting children write this word less often than it occurs in children's reading, and less than it is used by adults.

Table 5.6

Top 20 Most Common Adverbs In The OCCW (2020), OCCR (2019), BNC (Written Imaginative), And OEC. Asterisks Indicate Adverbs Unique To Being In The Top 20 Of That Corpus

	OCCW		OCCR		BNC		OEC	
#	Item	%	Item		Item	%	Item	%
1	up	0.43%	up	0.31%	not	1.23%	as	0.20%
2	out	0.30%	as	0.24%	so	0.30%	up	0.19%
3	then	0.28%	so	0.23%	then	0.27%	also	0.16%
4	back	0.26%	out	0.23%	now	0.21%	out	0.15%
5	just	0.22%	then	0.20%	just	0.19%	just	0.14%
6	so	0.22%	now	0.17%	back	0.18%	so	0.14%
7	down	0.21%	down	0.15%	as	0.18%	more	0.13%
8	as	0.19%	back	0.15%	well	0.14%	now	0.13%
9	there	0.19%	just	0.15%	there	0.13%	well	0.12%
10	now	0.14%	very	0.14%	here	0.13%	then	0.11%
11	very	0.14%	only	0.11%	too	0.13%	only	0.11%
12	away	0.12%	well	0.11%	again	0.12%	very	0.10%
13	off*	0.12%	away	0.11%	only	0.12%	even	0.10%
14	again	0.11%	again	0.10%	very	0.12%	here	0.08%
15	home*	0.10%	there	0.10%	never	0.12%	back	0.07%
16	never	0.10%	more	0.10%	even	0.12%	still	0.07%
17	here	0.09%	too	0.10%	still	0.11%	most	0.06%
18	really	0.09%	here	0.10%	all	0.10%	down	0.06%
19	still	0.09%	never	0.10%	away	0.09%	really	0.06%
20	only	0.09%	even	0.10%	more	0.09%	too	0.06%

5.2.6.5 Proper Nouns and Pronouns

The analysis of this part of speech presented some methodological problems with the tagging systems. The Oxford Children's Corpus collapses proper nouns and pronouns together, whereas the BNC does not distinguish between nouns and proper nouns, and so only has a pronoun

part of speech tag. The OEC, meanwhile, has a proper noun tag but no pronoun tag. Thus, only a few key facts are discussed here rather than a full frequency and keyword analysis.

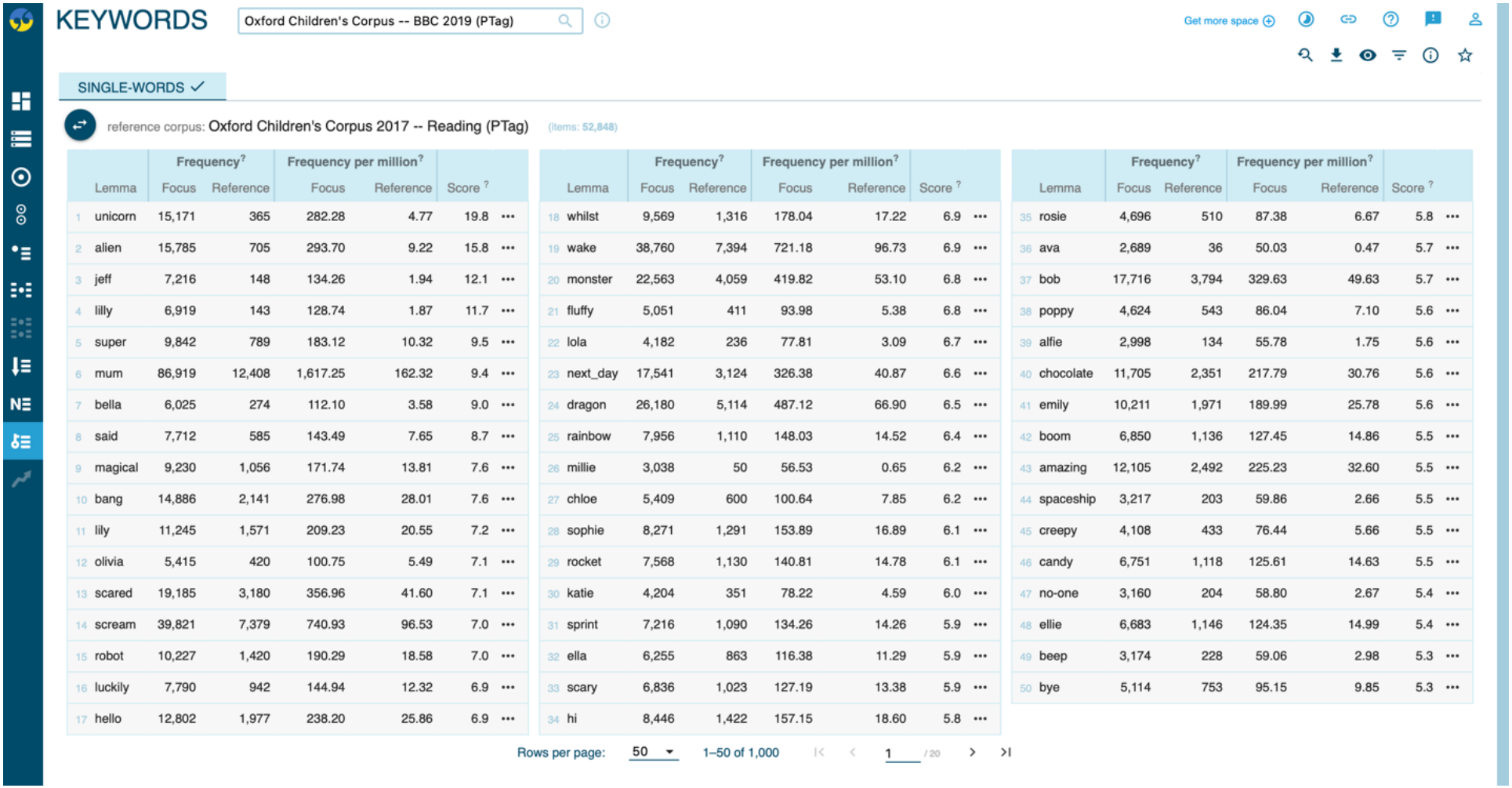
I, the first-person singular pronoun, appears over 1 million times in the OCCW corpus and is 2.21% of the entire corpus, the sixth most common word overall. In contrast, *I* occurs half as many times in the OCCR (0.86%), accounts for 1.75% of the BNC (1.75%), and only accounts for 0.01% of the OEC. This suggests that children write in the first person twice as much as texts written for children do and more than general adult fiction. This is likely due to the OCCR and OEC including non-fiction, or other texts that do not employ the first person.

5.2.7 Keywords and Collocation Analysis

Keyword comparisons between the OCCW and the OCCR, BNC, and OEC offer further insights into the differences between these corpora. In order to investigate the context of some of the words that are salient in the OCCW, the following sections present concordance and word sketch information for high-frequency words and keywords in the OCCW. These are organised by parts of speech, following the order of the preceding sections. First, I present examples of the results of the most basic keyword analyses using each reference corpus in turn. These offer a starting point for choosing the most salient words to investigate.

Figure 5.14

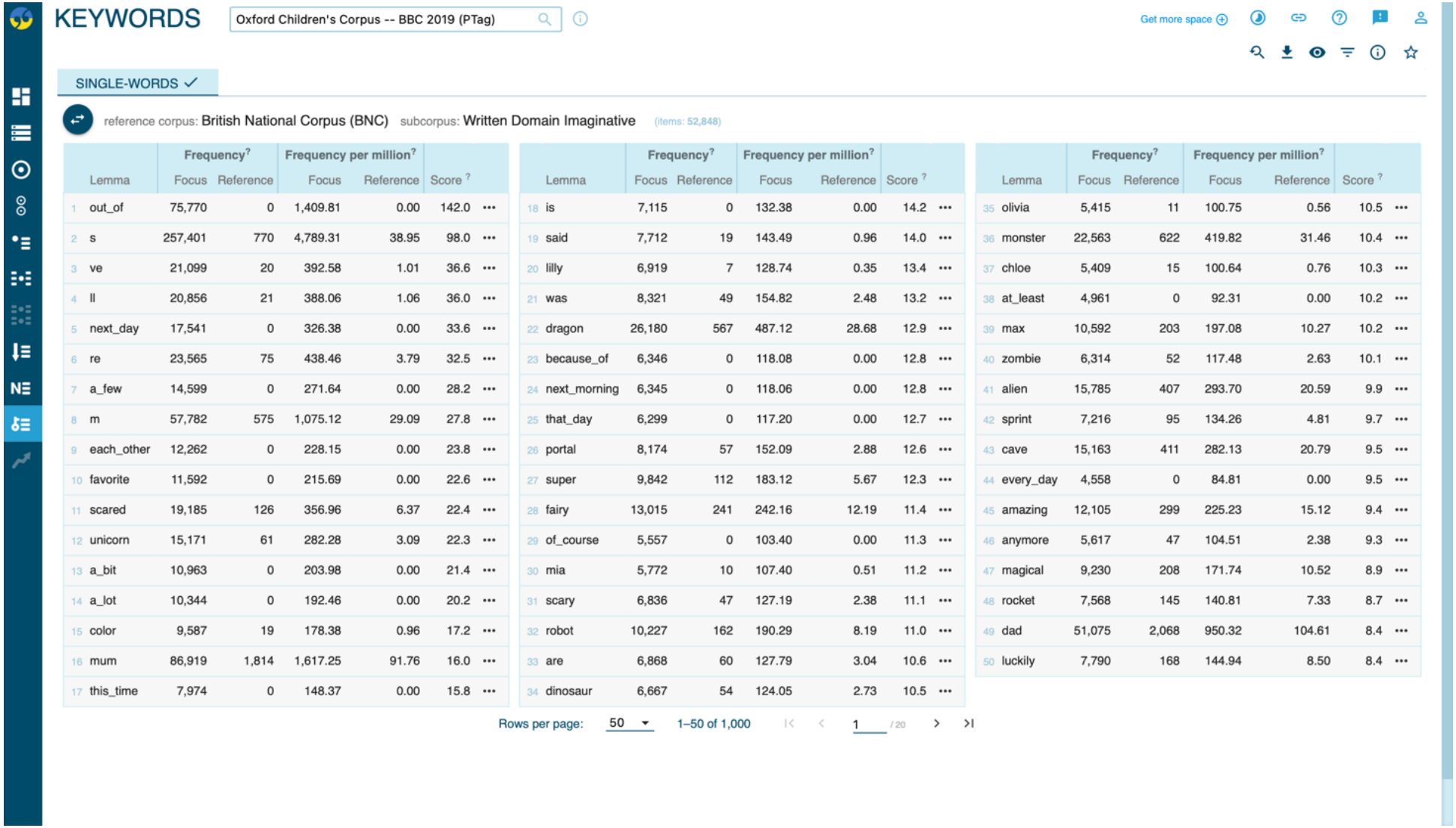
Keywords Using the OCCR as a Reference Corpus



	Lemma	Frequency Focus	Frequency Reference	Frequency per million Focus	Frequency per million Reference	Score
1	unicorn	15,171	365	282.28	4.77	19.8
2	alien	15,785	705	293.70	9.22	15.8
3	jeff	7,216	148	134.26	1.94	12.1
4	lilly	6,919	143	128.74	1.87	11.7
5	super	9,842	789	183.12	10.32	9.5
6	mum	86,919	12,408	1,617.25	162.32	9.4
7	bella	6,025	274	112.10	3.58	9.0
8	said	7,712	585	143.49	7.65	8.7
9	magical	9,230	1,056	171.74	13.81	7.6
10	bang	14,886	2,141	276.98	28.01	7.6
11	lily	11,245	1,571	209.23	20.55	7.2
12	olivia	5,415	420	100.75	5.49	7.1
13	scared	19,185	3,180	356.96	41.60	7.1
14	scream	39,821	7,379	740.93	96.53	7.0
15	robot	10,227	1,420	190.29	18.58	7.0
16	luckily	7,790	942	144.94	12.32	6.9
17	hello	12,802	1,977	238.20	25.86	6.9
18	whilst	9,569	1,316	178.04	17.22	6.9
19	wake	38,760	7,394	721.18	96.73	6.9
20	monster	22,563	4,059	419.82	53.10	6.8
21	fluffy	5,051	411	93.98	5.38	6.8
22	lola	4,182	236	77.81	3.09	6.7
23	next_day	17,541	3,124	326.38	40.87	6.6
24	dragon	26,180	5,114	487.12	66.90	6.5
25	rainbow	7,956	1,110	148.03	14.52	6.4
26	millie	3,038	50	56.53	0.65	6.2
27	chloe	5,409	600	100.64	7.85	6.2
28	sophie	8,271	1,291	153.89	16.89	6.1
29	rocket	7,568	1,130	140.81	14.78	6.1
30	katie	4,204	351	78.22	4.59	6.0
31	sprint	7,216	1,090	134.26	14.26	5.9
32	ella	6,255	863	116.38	11.29	5.9
33	scary	6,836	1,023	127.19	13.38	5.9
34	hi	8,446	1,422	157.15	18.60	5.8
35	rosie	4,696	510	87.38	6.67	5.8
36	ava	2,689	36	50.03	0.47	5.7
37	bob	17,716	3,794	329.63	49.63	5.7
38	poppy	4,624	543	86.04	7.10	5.6
39	alfie	2,998	134	55.78	1.75	5.6
40	chocolate	11,705	2,351	217.79	30.76	5.6
41	emily	10,211	1,971	189.99	25.78	5.6
42	boom	6,850	1,136	127.45	14.86	5.5
43	amazing	12,105	2,492	225.23	32.60	5.5
44	spaceship	3,217	203	59.86	2.66	5.5
45	creepy	4,108	433	76.44	5.66	5.5
46	candy	6,751	1,118	125.61	14.63	5.5
47	no-one	3,160	204	58.80	2.67	5.4
48	ellie	6,683	1,146	124.35	14.99	5.4
49	beep	3,174	228	59.06	2.98	5.3
50	bye	5,114	753	95.15	9.85	5.3

Figure 5.15

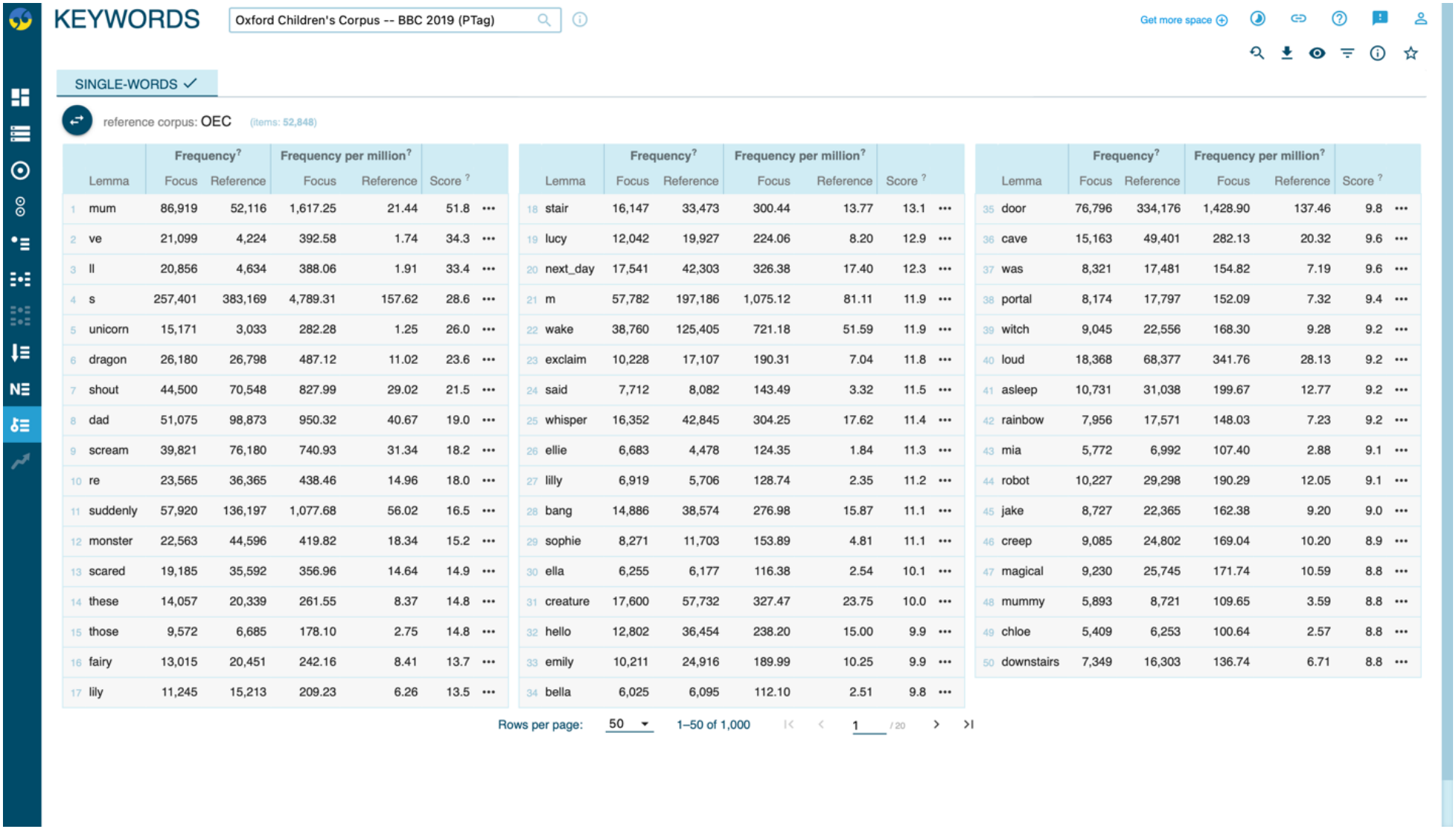
Keywords Using the BNC (Written Imaginative) as the Reference Corpus



	Lemma	Frequency Focus	Frequency Reference	Frequency per million Focus	Frequency per million Reference	Score
1	out_of	75,770	0	1,409.81	0.00	142.0
2	s	257,401	770	4,789.31	38.95	98.0
3	ve	21,099	20	392.58	1.01	36.6
4	ll	20,856	21	388.06	1.06	36.0
5	next_day	17,541	0	326.38	0.00	33.6
6	re	23,565	75	438.46	3.79	32.5
7	a_few	14,599	0	271.64	0.00	28.2
8	m	57,782	575	1,075.12	29.09	27.8
9	each_other	12,262	0	228.15	0.00	23.8
10	favorite	11,592	0	215.69	0.00	22.6
11	scared	19,185	126	356.96	6.37	22.4
12	unicorn	15,171	61	282.28	3.09	22.3
13	a_bit	10,963	0	203.98	0.00	21.4
14	a_lot	10,344	0	192.46	0.00	20.2
15	color	9,587	19	178.38	0.96	17.2
16	mum	86,919	1,814	1,617.25	91.76	16.0
17	this_time	7,974	0	148.37	0.00	15.8
18	is	7,115	0	132.38	0.00	14.2
19	said	7,712	19	143.49	0.96	14.0
20	lilly	6,919	7	128.74	0.35	13.4
21	was	8,321	49	154.82	2.48	13.2
22	dragon	26,180	567	487.12	28.68	12.9
23	because_of	6,346	0	118.08	0.00	12.8
24	next_morning	6,345	0	118.06	0.00	12.8
25	that_day	6,299	0	117.20	0.00	12.7
26	portal	8,174	57	152.09	2.88	12.6
27	super	9,842	112	183.12	5.67	12.3
28	fairy	13,015	241	242.16	12.19	11.4
29	of_course	5,557	0	103.40	0.00	11.3
30	mia	5,772	10	107.40	0.51	11.2
31	scary	6,836	47	127.19	2.38	11.1
32	robot	10,227	162	190.29	8.19	11.0
33	are	6,868	60	127.79	3.04	10.6
34	dinosaur	6,667	54	124.05	2.73	10.5
35	olivia	5,415	11	100.75	0.56	10.5
36	monster	22,563	622	419.82	31.46	10.4
37	chloe	5,409	15	100.64	0.76	10.3
38	at_least	4,961	0	92.31	0.00	10.2
39	max	10,592	203	197.08	10.27	10.2
40	zombie	6,314	52	117.48	2.63	10.1
41	alien	15,785	407	293.70	20.59	9.9
42	sprint	7,216	95	134.26	4.81	9.7
43	cave	15,163	411	282.13	20.79	9.5
44	every_day	4,558	0	84.81	0.00	9.5
45	amazing	12,105	299	225.23	15.12	9.4
46	anymore	5,617	47	104.51	2.38	9.3
47	magical	9,230	208	171.74	10.52	8.9
48	rocket	7,568	145	140.81	7.33	8.7
49	dad	51,075	2,068	950.32	104.61	8.4
50	luckily	7,790	168	144.94	8.50	8.4

Figure 5.16

Keywords Using the OEC as the Reference Corpus



	Lemma	Frequency Focus	Frequency Reference	Frequency per million Focus	Frequency per million Reference	Score
1	mum	86,919	52,116	1,617.25	21.44	51.8
2	ve	21,099	4,224	392.58	1.74	34.3
3	ll	20,856	4,634	388.06	1.91	33.4
4	s	257,401	383,169	4,789.31	157.62	28.6
5	unicorn	15,171	3,033	282.28	1.25	26.0
6	dragon	26,180	26,798	487.12	11.02	23.6
7	shout	44,500	70,548	827.99	29.02	21.5
8	dad	51,075	98,873	950.32	40.67	19.0
9	scream	39,821	76,180	740.93	31.34	18.2
10	re	23,565	36,365	438.46	14.96	18.0
11	suddenly	57,920	136,197	1,077.68	56.02	16.5
12	monster	22,563	44,596	419.82	18.34	15.2
13	scared	19,185	35,592	356.96	14.64	14.9
14	these	14,057	20,339	261.55	8.37	14.8
15	those	9,572	6,685	178.10	2.75	14.8
16	fairy	13,015	20,451	242.16	8.41	13.7
17	lily	11,245	15,213	209.23	6.26	13.5
18	stair	16,147	33,473	300.44	13.77	13.1
19	lucy	12,042	19,927	224.06	8.20	12.9
20	next_day	17,541	42,303	326.38	17.40	12.3
21	m	57,782	197,186	1,075.12	81.11	11.9
22	wake	38,760	125,405	721.18	51.59	11.9
23	exclaim	10,228	17,107	190.31	7.04	11.8
24	said	7,712	8,082	143.49	3.32	11.5
25	whisper	16,352	42,845	304.25	17.62	11.4
26	ellie	6,683	4,478	124.35	1.84	11.3
27	lilly	6,919	5,706	128.74	2.35	11.2
28	bang	14,886	38,574	276.98	15.87	11.1
29	sophie	8,271	11,703	153.89	4.81	11.1
30	ella	6,255	6,177	116.38	2.54	10.1
31	creature	17,600	57,732	327.47	23.75	10.0
32	hello	12,802	36,454	238.20	15.00	9.9
33	emily	10,211	24,916	189.99	10.25	9.9
34	bella	6,025	6,095	112.10	2.51	9.8
35	door	76,796	334,176	1,428.90	137.46	9.8
36	cave	15,163	49,401	282.13	20.32	9.6
37	was	8,321	17,481	154.82	7.19	9.6
38	portal	8,174	17,797	152.09	7.32	9.4
39	witch	9,045	22,556	168.30	9.28	9.2
40	loud	18,368	68,377	341.76	28.13	9.2
41	asleep	10,731	31,038	199.67	12.77	9.2
42	rainbow	7,956	17,571	148.03	7.23	9.2
43	mia	5,772	6,992	107.40	2.88	9.1
44	robot	10,227	29,298	190.29	12.05	9.1
45	jake	8,727	22,365	162.38	9.20	9.0
46	creep	9,085	24,802	169.04	10.20	8.9
47	magical	9,230	25,745	171.74	10.59	8.8
48	mummy	5,893	8,721	109.65	3.59	8.8
49	chloe	5,409	6,253	100.64	2.57	8.8
50	downstairs	7,349	16,303	136.74	6.71	8.8

Some general conclusions about the keywords in the OCCW can be drawn by considering these results. It is plain that this sample of children's writing is characterised by references to fantastical and fictional creatures and some particular adjectives, adverbs and verbs indicative of narrative texts, as well as a large proportion of proper nouns (names of characters). Some keywords are now examined in more detail in terms of separate parts of speech.

5.2.7.1 Nouns

Ignoring proper nouns (i.e., character names), the key nouns in the OCCW are largely related to fictional creatures (*unicorn, dragon, monster, fairy, alien, zombie, mummy, witch*), places (*cave, planet*), creatures (*dinosaur, shark, monkey, puppy*), and objects (*portal, potion, rocket, robot*). Many of these key nouns do not appear when using the OCCR as the reference corpus, indicating that these nouns are used in similar amounts in both the OCCR and OCCW, and thus that children's writing and children's reading are thematically similar. These data support the findings of prior corpus-based research into children's literature which found core vocabulary common in child-directed texts (Dawson et al., 2021; Thompson & Sealey, 2007; Wild et al., 2013).

These themes are in obvious contrast to writing produced by and for adults. The OEC's key nouns have a focus on words related to current affairs (*policy, development, authority, industry, government, programme*). As for the BNC, its key nouns represent themes of intimacy, relationships, and other themes found in adult fiction: *breast, marriage, pleasure, affair, bastard, cigarette, sex, telephone, doubt, shrug, wine*. Obviously, these are not going to be present in children's fiction and they are also not common in children's writing.

Thus, we can draw the conclusion that children employ specific low-frequency vocabulary that does not occur to the same extent in other English corpora. The words occurring in the OCCW more than in the other corpora represent semantic areas that are particularly relevant to children, such as people, places, and nature. Their writing shows children's interest in fictional creatures, monsters, animals, and magical objects.

5.2.7.2 Verbs

Several key verbs characterise this sample of children's writing, such as those related to speech, e.g., *shout*, *scream*, *whisper, exclaim, say, cry, reply*; actions such as *wake*, *grab*, *stare*; and others, e.g., *creak*. These are indicative of key features of children's fiction, such as direct speech acts and emotive actions (Stephens, 2005). *Go* is far more common in the OCCW (0.81%) than it is in the OCCR (0.36%), BNC (0.40%), or OEC (0.21%). Due to children's focus on physical space, direct speech acts, and movement through worlds, as well as expressive language close to the self (Britton, 1972), there is a reliance on core verbs to express these actions (Thompson & Sealey, 2007; Wild et al., 2013).

5.2.7.3 Adjectives

A keyword analysis reveals key adjectives using the OCCW corpora as a focus corpus and the OCCR, BNC and OEC as reference corpora. *Scared/scary, super, favourite, magical, excited, fluffy, amazing*, and *weird* are adjectives which characterise this sample of children's writing. Other key adjectives are *cute, worried*, and *rainbow* (used as part of a compound adjective i.e., rainbow-colored). In comparison, there are few key adjectives in the BNC except *grey, closer*, and *least*; this suggests that the BNC gets by with the 'core' adjectives, many of which are also frequent in the OCCW. In contrast, the OEC has several adjective keywords that do not appear in the other corpora, mainly related to current affairs, such as *economic, political, available*, and *financial*. These are unsurprisingly serious, "adult" topics that do not occur in other sources of writing and represent the different thematic areas that children and adults address in their writing.

5.2.8 Age Effects

To answer "RQ3: To what extent does the writing of the 5–9 group differ from the writing of the 10–13 group?", the corpora were investigated for any trends and effects in relation to age. Firstly, because the use of different parts of speech has been shown to change as children age

(Durrant and Brenchley, 2019), the use of parts of speech were investigated for each age group in the OCCW.

Figure 5.17

Proportions of Parts Of Speech By Age In The OCCW

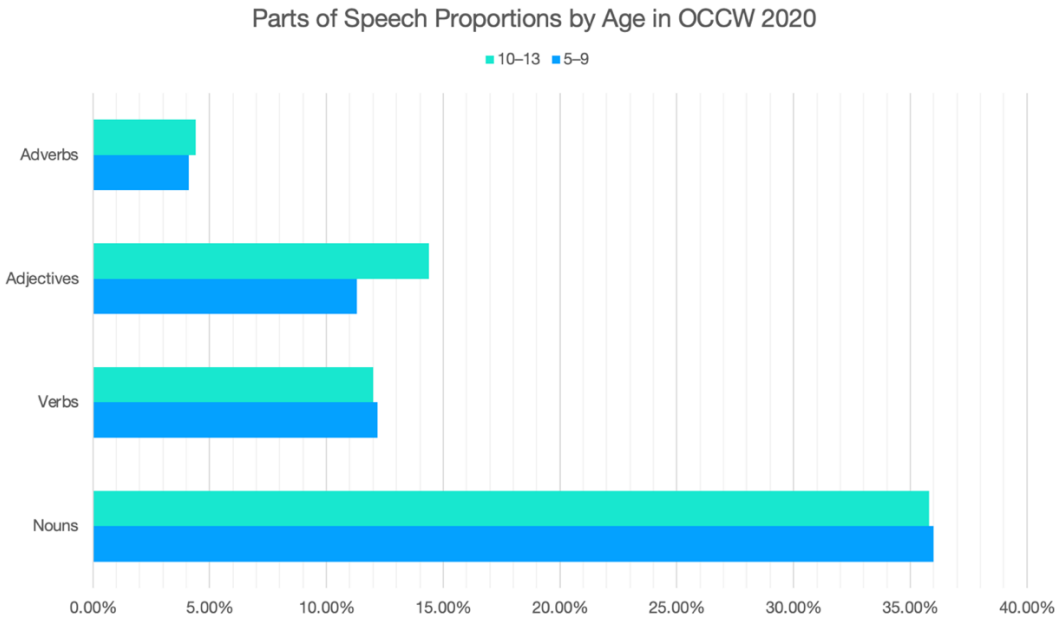


Table 5.7

Proportions of Parts of Speech By Age In The OCCW

Age	All lemmas	Nouns	Verbs	Adjectives	Adverbs
5–9	30,611	11,014 (36.0%)	3736 (12.2%)	3453 (11.3%)	1242 (4.1%)
10–13	38,033	13,602 (35.8%)	4546 (12.0%)	5484 (14.4%)	1681 (4.4%)

The results showed that the distributions of parts of speech were relatively similar, but that the older group used more adjectives and adverbs compared to the younger group. The younger group used slightly more nouns and verbs than the older group. The next goal was to examine what words were typical in each group. To facilitate this, the 5–9 sub-corpus and 10–13 sub-corpus were

used as reference/focus corpora between each other using Sketch Engine's keyword analysis feature. Simple n = 11, minimum frequency = 5.

As we start to conduct the analysis, the limitations of this method are revealed. Using Sketch Engine's default keyword 'focus on' number (1), leads to the following screens in Figure 5.16 and 5.17. We can see misspellings and proper nouns (e.g., names) end up in the keyword list. The screens show that the corpus has not been corrected for spelling (e.g., *adrenalin* rather than *adrenaline*, the first word on the older writers' list). This means that misspellings will cloud the results we are interested in (but does have the advantage of being able to investigate misspellings if interested).

Figure 5.18

Words More Common in The 5–9 Sub-Corpus Compared To The 10–13 Sub-Corpus (OCCW 2020) Organised By Relative Frequency

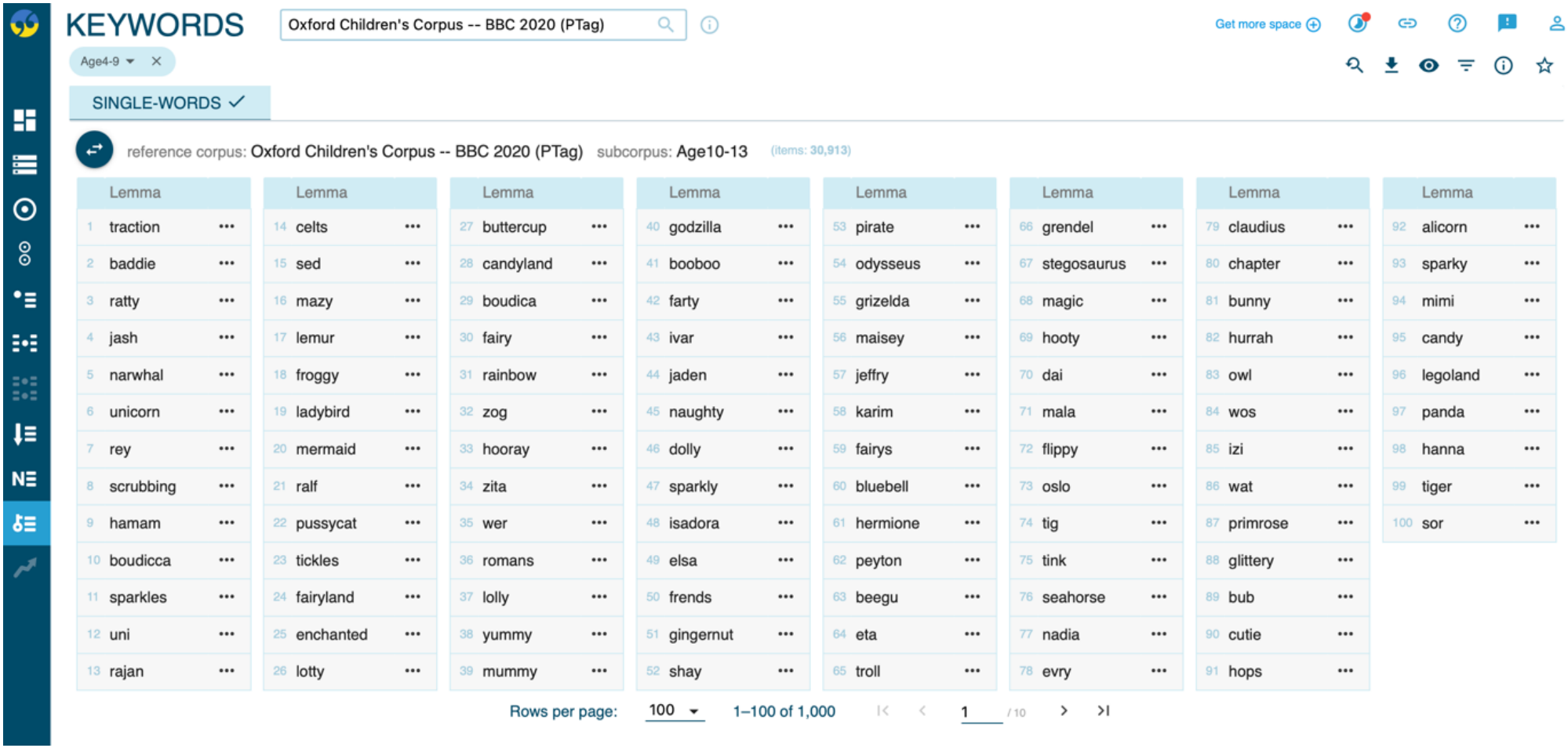
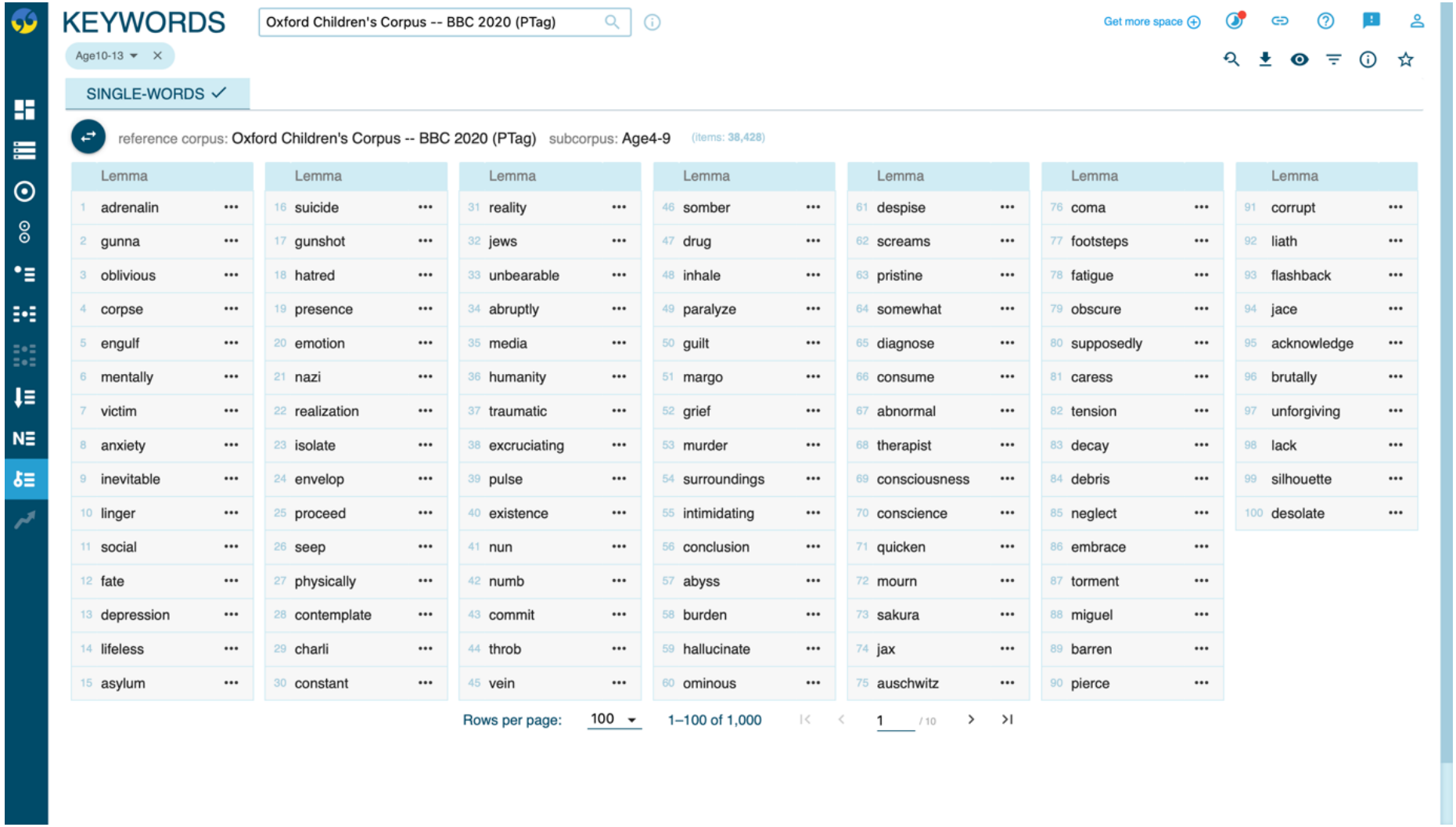


Figure 5.19

Words More Common in The 10–13 Sub-Corpus Compared To The 5–9 Sub-Corpus (OCCW 2020) Organised By Relative Frequency



Adjusting the keyword number to 10, i.e., a focus on 'more common' words, eliminates the majority of these misspellings and proper nouns. The top 50 results for each corpus are shown in Table 5.9, allowing for a thematic analysis of keywords in this sample of children's writing, dependent on age. These are organised by keyness score, as calculated by Sketch Engine, which is the default way to order results. Table 5.10 illustrates the raw frequencies for the top 30 for each corpus.

Table 5.8

Top 50 Keywords From The Oxford Children's Corpora Depending On Age

Younger (4–9)		**Older (10–13)**	
unicorn	volcano	pain	pale
fairy	daddy	victim	throat
mermaid	wand	murder	adrenalin
rainbow	super	reality	glance
mummy	romans	emotion	thought
magic	pearl	lifeless	sense
chapter	sparkle	engulf	memory
pirate	robber	pierce	attempt
tiger	marshmallow	silence	tear
princess	elf	fate	death
owl	potion	surroundings	isolate
candy	dolphin	soul	blood
dragon	enchant	hell	vein
yummy	poppy	vision	social
witch	crocodile	myself	torture
bunny	castle	sweat	anxiety
naughty	baddie	lung	pace
troll	treasure	abruptly	trickle
magical	santa	silhouette	proceed
sparkly	ninja	innocent	corpse
wizard	lot	comfort	due
rosie*	queen	gunshot	suffer
evie*	happily	grasp	presence
panda	daisy	regret	tears
lilly*	penguin	spine	rot

Table 5.9

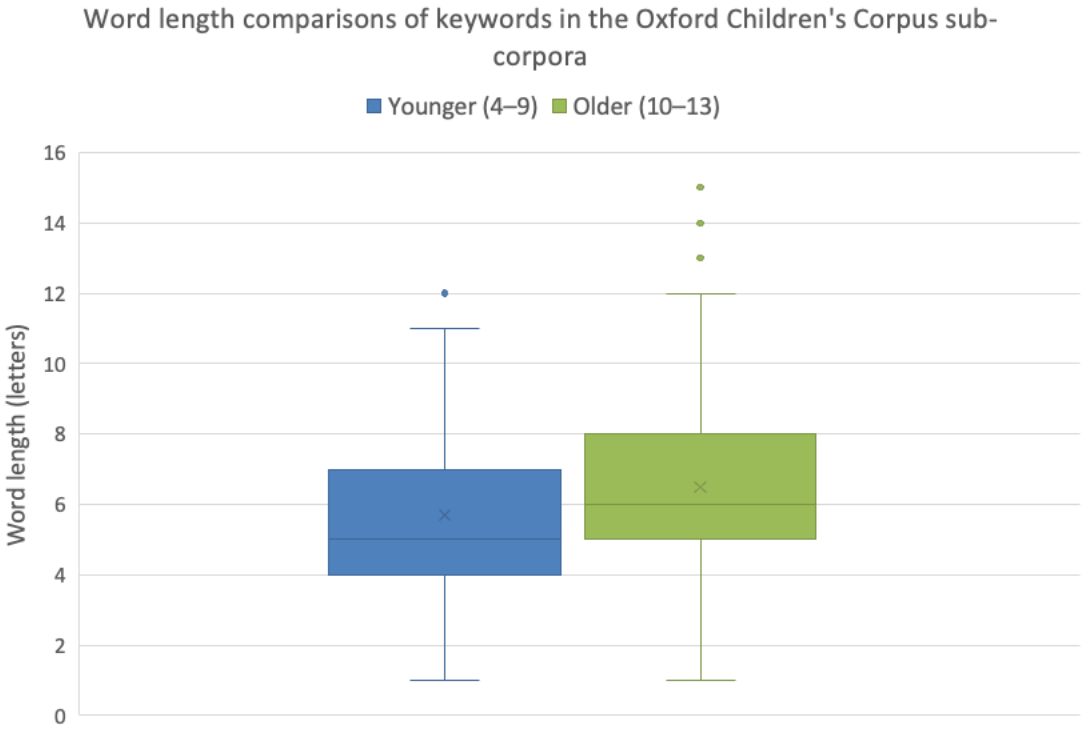
Frequencies for Keywords in the Oxford Children's Corpus by Age

Younger (4–9)			Older (10–13)		
Item	**Frequency (focus)**	**Frequency (reference)**	**Item**	**Frequency (focus)**	**Frequency (reference)**
unicorn	7264	2249	*pain*	6348	1306
fairy	7781	3208	*victim*	1378	178
mermaid	2692	986	*murder*	2581	453
rainbow	4464	1838	*reality*	1699	274
mummy	4634	1947	*emotion*	1360	196
magic	8856	4324	*lifeless*	1176	157
chapter	3824	1891	*engulf*	1008	123
pirate	2216	1030	*pierce*	2011	375
tiger	3914	2038	*silence*	5630	1298
princess	4477	2377	*fate*	1017	133
owl	2412	1200	*surroundings*	1453	249
candy	3000	1552	*soul*	2733	597
dragon	13405	7684	*hell*	1536	287
yummy	1111	447	*vision*	1821	368
witch	5183	2893	*myself*	13355	3539
bunny	1833	895	*sweat*	2942	692
naughty	1197	524	*lung*	1415	275
troll	1448	685	*abruptly*	878	135
magical	5631	3311	*silhouette*	1116	204
sparkly	1140	508	*innocent*	1641	355
wizard	2831	1573	*comfort*	1566	335
rosie	3212	1827	*gunshot*	717	91
evie	1794	945	*grasp*	1400	289
panda	1576	806	*regret*	1408	292
lilly	4812	2875	*spine*	2609	640
volcano	2195	1205	*pale*	2889	722
daddy	2180	1202	*throat*	1679	376
wand	1475	760	*adrenalin*	572	53
super	5775	3662	*glance*	2863	723
romans	685	265	*thought*	7079	1966
pearl	1237	643	*sense*	3492	915

In terms of theme, the younger writers use more nouns related to fantastical or mythical beasts, such as *unicorn* and *mermaid*, as well as animals in general, such as *froggy* and *lemur*. In comparison, the older writer corpus is characterised by longer and more complex words, such as *excruciating* or *consciousness*. Many of these words employed by the older writers seem to generally deal with darker themes, such as *suicide*, *anxiety*, *depression*, *grief* and *murder*. The older writers use generally longer words (6.49 letters versus 5.7, on average, taking the first 1,000 keywords; n=10, min.freq=5). If we use word length as a proxy for sophistication, it suggests older writers are using more sophisticated vocabulary.

Figure 5.20

Word Lengths In Keywords In The Oxford Children's Corpus By Age



Keywords only capture a small amount of lexical information and are decontextualised, however. Further analysis of the differences in writing between the older and younger groups

requires being able to manipulate and analyse the data in context, in full, preferably offline (to allow for flexibility and the opportunity to use ready-made or bespoke qualitative corpus analysis programmes). Unfortunately, I was not able to access the corpus in a format that would facilitate this. The next section presents a method to further analyse the writing in this corpus for age effects.

5.3 Study 2: Mini-Corpus

The second arm of this chapter aimed to more deeply investigate a specific subset of stories submitted to the 500 Words competition to answer RQ2 (*What characterises the writing of the winners of the short-story competition? To what extent do the winners' stories differ from the other entries?)* and also to contribute information for RQ3 (*To what extent does the writing of the 5–9-year-old group differ from the writing of the 10–13-year-olds group?*). However, this section does not provide an answer to RQ2, due to pandemic restrictions limiting access to the corpus held by Oxford University Press. Instead, I used a smaller corpus constructed from the winners' entries to provide more answers to RQ3, in addition to the age-related differences presented above. I will now explain the limitations with RQ2, then proceed to detail what I was able to gather related to the differences in lexical features of written language between the two age groups who submit to the BBC 500 Words competition.

When children submit entries to the competition, they are distributed to volunteer judges who receive a randomly allocated batch of around 20 stories. These are then marked by each judge on a rubric with criteria relating to quality; the winners of each year are chosen based on this rubric. Therefore, creating a corpus with the winners' stories would technically allow for the analysis of a subset of the stories submitted to the competition more deeply and in relation to one common axis used to analyse writing development, i.e., quality. This corpus could thus be judged as a peer target

corpus, in that this is the standard that entrants to the competition are hoping to reach (in contrast to the prospective target corpora used in Study 1).

If we wanted to compare the winners to the rest of this corpus based on this rubric, we would first need to understand the rubric in detail. The current rubric is explained in a little detail here: https://www.bbc.co.uk/teach/500-words/criteria/zsmh3qt (accessed 1 August 2024), showing that stories are marked on 'Originality', 'Characterisation', 'Plot', 'Language', and 'Enjoyment'. The detailed rubric was not available, however. We would also need to know the ratings for all the stories in the entire corpus; however, this information was not available either. Comparing the winners based on this axis was thus deemed infeasible.

Another way to compare the quality would be to statistically examine quantifiable lexical features as discussed in Chapter 2.4. We could examine all the entries and compare them to the winners, to see if there are any lexical features that separate the winners from the rest of the corpus. However, examining these quantifiable vocabulary indices is a task which requires offline files (usually in .txt format), and this data was not available for the corpus as a whole, as it is held in Sketch Engine. Further complicating matters was that the winners' entries were not separated or tagged as such in Sketch Engine, either, or available as a separate corpus.

It was thus apparent that the two datasets were not available in comparable formats (the entire corpus held in Sketch Engine only, and the winners' corpus held in offline .txt files). Thus, it was evident that very little work could be done to answer RQ2. However, as the winners are published in age groups (the 5–9-years-old age group and the 10–13-years-old age group), we can use the corpus to investigate age-related differences, i.e., to answer RQ3 (*To what extent does the writing of the 5–9-year-old group differ from the writing of the 10–13-year-olds group?).*

The winning stories were copied and pasted from the BBC 500 Words website into separate .txt files, separated by age group, to create a corpus. Data from the last three available years (2020,

2019, 2018) were used to create a corpus of 81,414 words. There were 83 stories from the 5–9 group and 85 from the 10–13 group, for a total of 168 narrative texts. As highlighted by previous research (Durrant, 2022a; Durrant et al., 2021), the first step in research on lexical richness is to carefully define the constructs to be measured. The following features were identified as the object of investigation: word frequency, word range, academic language, concreteness, and lexical diversity. The indices used to evaluate these texts are summarised in Table 5.12 alongside a summary for their justification and the software used to analyse them. The following section defines these constructs more clearly and descriptions of the software used and methodological procedures are given.

Table 5.12

Indices Used To Evaluate The Winners' Corpus

Construct	Measure	Software	Description	References
Lexical sophistication	Word frequency	TAALES (COCA_fiction_Frequency_Log_AW and BNC_Written_Freq_AW_Log)	Using logarithmic frequency information from COCA (fiction) and BNC.	Kyle et al., 2018; Kyle & Crossley, 2015
	Word range	TAALES (BNC_Written_Range_AW)	Using range indices based on the written subset of the BNC (3083 texts).	Kyle et al., 2018; Kyle & Crossley, 2015
	Academic language	TAALES	Using academic lists (AWL and AFL) to calculate frequency of academic vocabulary.	Kyle et al., 2018; Kyle & Crossley, 2015
	Concreteness	TAALES (Brysbaert_Concreteness_Combined_AW)	Using concreteness ratings from 37,058 lemmas.	Brysbaert et al., 2014
Lexical diversity	MTLD	TAALED (MTLD_Original_AW)	MTLD is based on the average number of tokens it takes to reach a given TTR value (.720).	Mccarthy, 2005; McCarthy & Jarvis, 2010; Kyle et al., 2021
	HD-D	TAALED (HD-D_AW)	HD-D uses the hypergeometric distribution to calculate the probability of encountering one of its tokens in a random sample of 42 tokens. For ease of interpretation, this is converted to the same scale as TTR.	McCarthy & Jarvis, 2007, 2010; Zenker & Kyle, 2021; Kyle et al., 2021

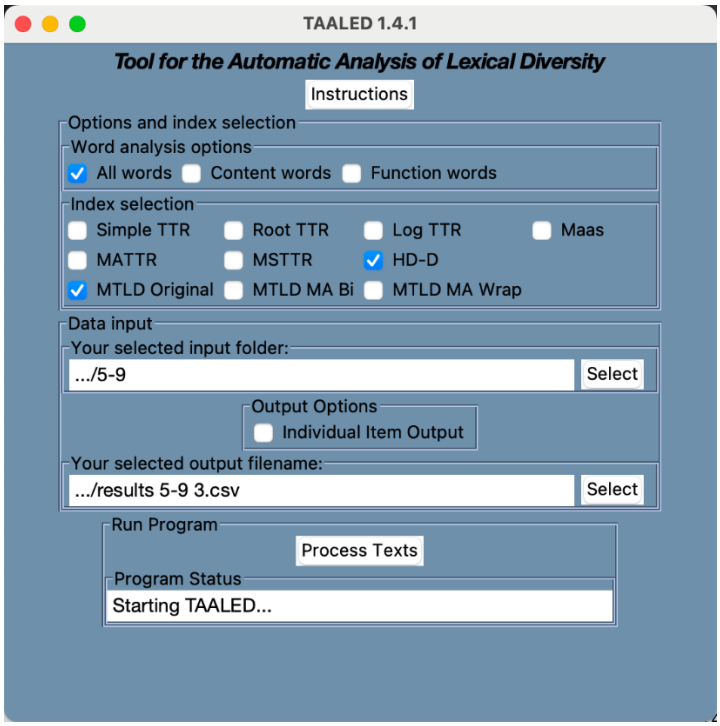
5.3.1 Lexical Diversity

Lexical diversity is the assessment of the amount of variation and repetition in a written or spoken text, with the assumptions that a more diverse text has more variation, less repetition, and so signals a higher quality of language. Lexical diversity has been linked to both evaluations of time and quality in L1 vocabulary research; for a full analysis of the research results on lexical diversity, please see Section 2.4 of the literature review. The remainder of this section summarises how the *methods* used to calculate lexical diversity have changed over the years.

In this study, the Tool for the Automatic Analysis of Lexical Diversity (TAALED) was employed to measure lexical diversity. TAALED calculates a range of classic and more robust indices of lexical diversity, e.g., TTR, MATTR and MTLD. TAALED is open-source and freely available to use. It uses lemmatised forms of words, and texts can be assessed for content words, function words, or all words. In this study, MTLD and HD-D were used. The former has been demonstrated to be independent of text length (Zenker & Kyle, 2021), and the latter is a more reliable calculation of vocd, which has been shown to have a negligible relationship with text length (McCarthy & Jarvis, 2010). In TAALED, HD-D scores are converted into the same log scale as their TTR calculations.

Figure 5.21

TAALED Execution Screen



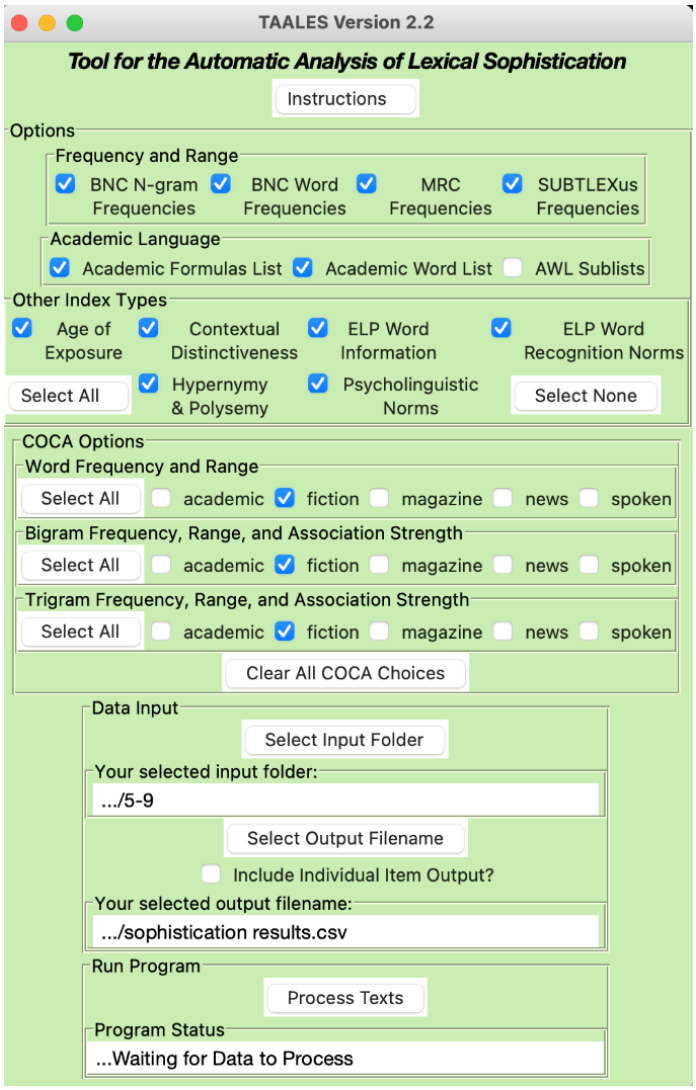
5.3.2 Lexical Sophistication

Lexical sophistication is an important idea in vocabulary studies and other fields of psychology and linguistics. Studies have operationalised lexical sophistication as rare words (Arnaud, 1984), difficult words (Meara & Bell, 2001) or low-frequency words (Linnarud, 1986; Read, 2000), and these ideas are all based on the assumption that learners with higher proficiencies and larger vocabularies are more likely to use more sophisticated, and less frequent, words (Meara & Bell, 2001). The use of rare words has been to linked to higher language proficiency (Arnaud, 1984; Linnarud, 1986) and research has shown that texts with more low-frequency vocabulary are rated as higher quality and difficulty (Kyle et al., 2020; D. S. McNamara et al., 2010). For a fuller analysis of the research results on lexical sophistication, please see Section 2.4.2 of the literature review.

In this study, the Tool for the Automatic Analysis of Lexical Sophistication 2.0 (Kyle et al., 2018; Kyle & Crossley, 2015) was employed to measure lexical sophistication. This is a freely available and open-source piece of software which produces a range of measures linked to lexical sophistication. The first version included 104 indices linked to word frequency, range, academic language, n-gram frequency, and psycholinguistic norms (e.g., abstractness, concreteness). The second version (used here) updated and refined these indices, and now TAALES can produce over 400 indices for lexical sophistication. Due to the wide range of indices, careful consideration is needed when deciding which to evaluate. Lexical sophistication measures commonly used in the literature include word frequency, register specificity, word length, and psycholinguistic features of words e.g., concreteness and familiarity (S. A. Crossley et al., 2011; Guo et al., 2013). TAALES contains indices covering these measures of lexical sophistication.

Figure 5.22

TAALES Execution Screen



I focused on three areas: (1) frequency; (2) range; and (3) concreteness. Word frequency was operationalised based on using the COCA fiction and BNC frequency results as reference corpora with log transformation applied to avoid the problem of extremely high-frequency words skewing results (i.e., the Zipfian effects common in word frequency lists). Range (also known as dispersion, entropy or contextual diversity) was also measured, under the assumption that words that are used in fewer contexts are judged to be more sophisticated. Range accounts for how widely a word is used (rather than how many times a word occurs in a corpus as a whole) by counting the

number of documents in which that word occurs. In TAALES, the BNC Written Range index has been calculated to explain 25.9% of the variance in holistic lexical proficiency scores, with range scores negatively correlated with proficiency scores (i.e., words used in fewer contexts are judged to be more sophisticated) (Kyle & Crossley, 2015). Concreteness is the idea that less frequent words are often more abstract and research has shown that, generally, children use more abstract words as they age (Nippold et al., 2005; L. Sun & Nippold, 2012). Concreteness was operationalised, via TAALES, using norms calculated by Brysbaert et al. (2014), who generated concreteness ratings for 40,000 English lemmas using online crowdsourcing data collection.

In all cases, TAALES calculates scores for all words in a text, then provides an average result for each overall text submitted into the program. These mean results were imported as a .csv file to Excel, and were then further processed.

5.3.3 Results

Firstly, the corpus was imported into Sketch Engine to see if the Sketch Engine tools could be used to illuminate any differences between the winners' corpus and the general submission corpus. However, keyword results were inconclusive, mainly bringing up proper nouns of main characters used in the winners' stories, even upon adjusting the keyword parameters. This was probably due to the size of the winners' corpus being small.

Analysis proceeded using the .txt files and TAALES/TAALED. The lexical diversity results are presented in Figure 5.20 and Figure 5.21 for HD-D and MTLD, respectively. In both, we can see a difference between the scores in the two age groups, with the older children scoring higher on lexical diversity than the younger years. The older group has a higher mean score in both indices (HD-D 0.82; MTLD 74.5) than the younger group (HD-D 0.80; MTLD 70.6). The differences in the means are very small and there are overlaps between the overall distributions (see graphs on the following pages), so we cannot make strong claims about the differences in year groups. This

increase in diversity with increasing age supports prior research associating measures of lexical diversity with age in L1 writers (R. A. Berman & Nir-Sagiv, 2004; S. A. Crossley et al., 2011; Malvern et al., 2004). We can see a large variation in the scores in each group, which is most likely due to the fact that several ages have been collapsed into two disparate groups, and there is, for example, a lot of difference in vocabulary knowledge between a 5-year-old and a 9-year-old. If each year of age could be measured separately, we would be able to see more distinct scores. Ideally, the stories would have been imported from Sketch Engine, which tags for age in the metadata. However, due to privacy regulations, Oxford University Press was not able to provide raw data of the type that is needed to conduct the analyses in this section.

Figure 5.23

HD-D Scores By Age Group In The Winners' Corpus

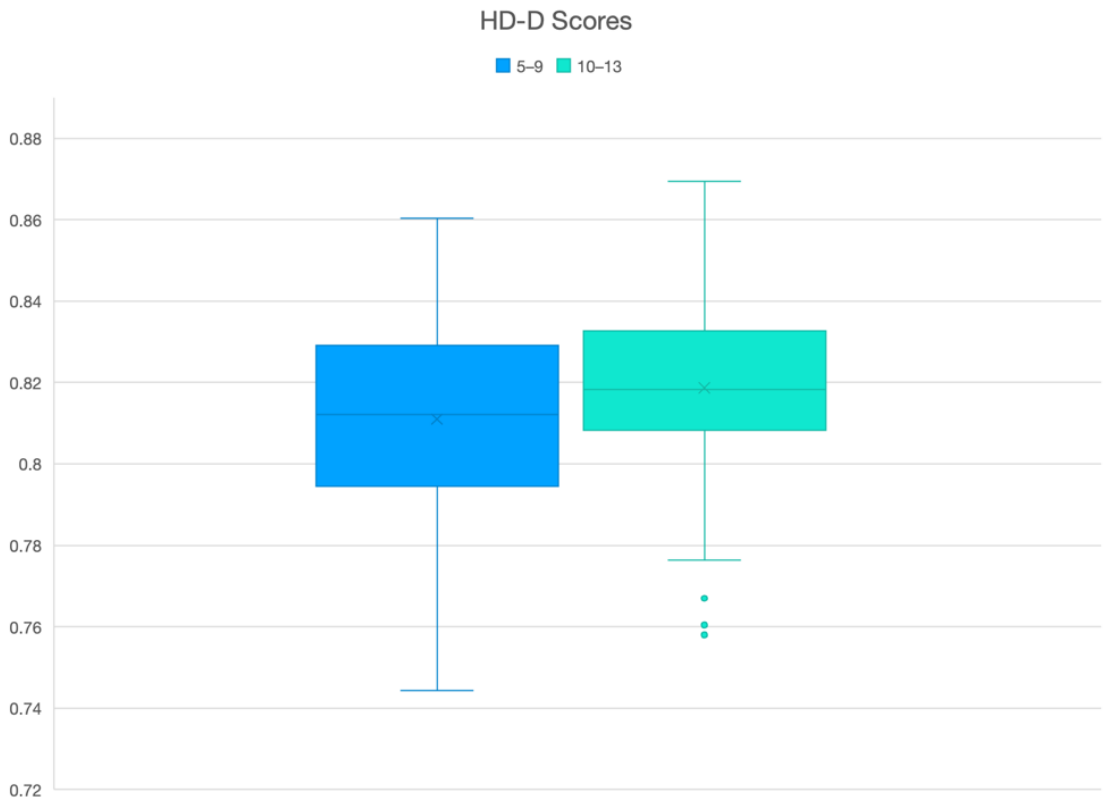
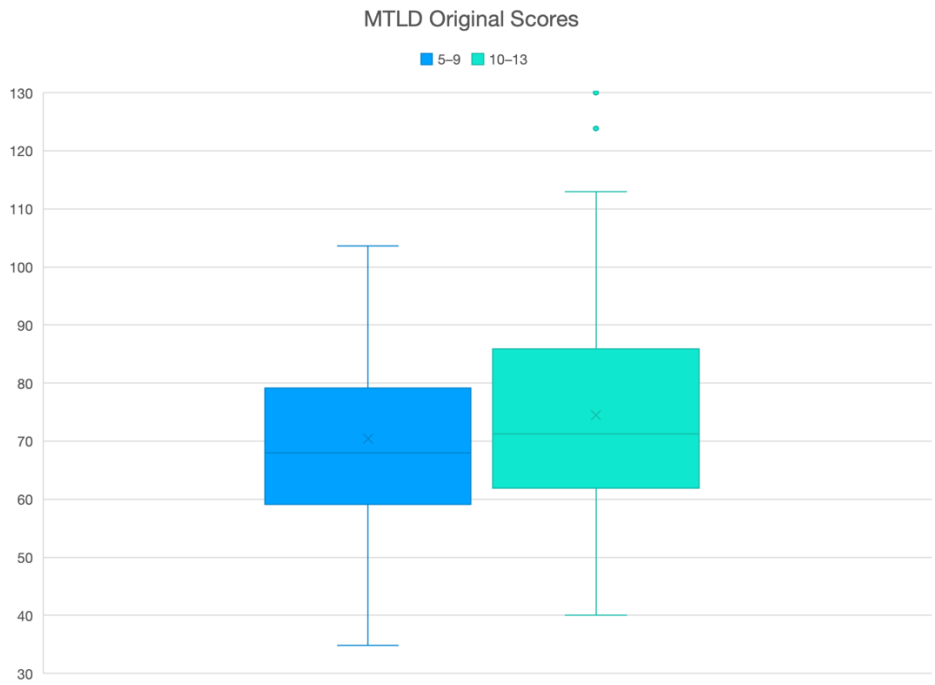


Figure 5.24

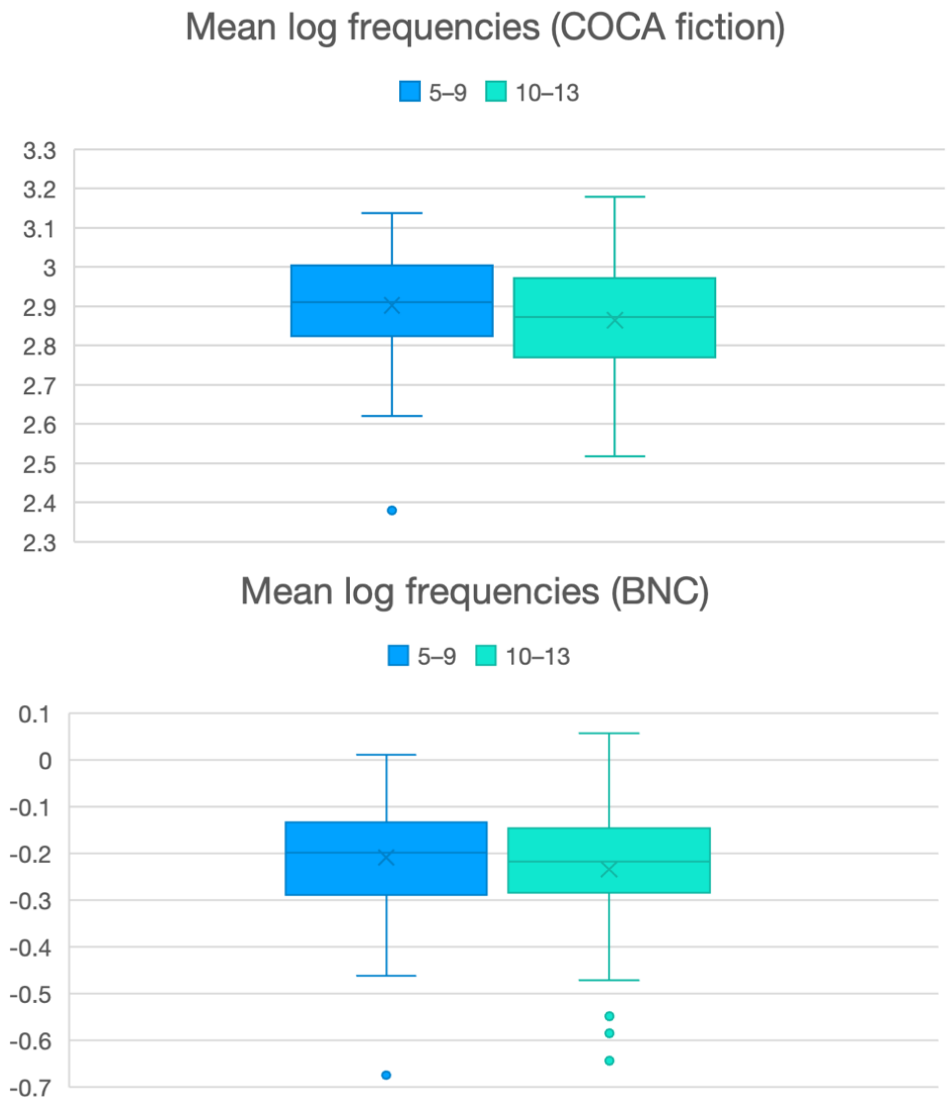
MTLD Scores By Age Group In The Winners' Corpus



Word frequency results showed that the older age group used more low-frequency vocabulary than the younger age group, supporting the idea that lexical sophistication increases with age (Olinghouse & Leaird, 2009; Olinghouse & Wilson, 2013; Y. Sun et al., 2010). Older writers also used a larger range of vocabulary. Again, these are descriptive results, and we cannot make any hard claims about the differences between the means or the distribution because the results have not been investigated for statistical significance. One reason for this is due to the overlap in ranges. Ideally, more fine-grained analysis linking frequency to part-of-speech would allow us to investigate whether there are specific variations in frequency by age in different grammatical categories, as was found by Durrant and Brenchley (2019); however, this was not possible using TAALES and further analysis with other methods was outside the limits of this investigation.

Figure 5.25

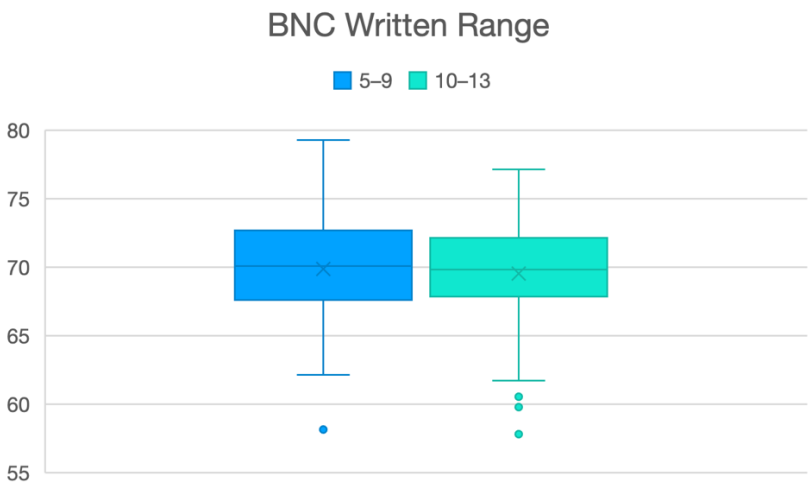
Mean Log Frequencies For All Words By Age Group In The Winners' Corpus



Furthermore, older writers also showed increased lexical sophistication when their texts were evaluated for range (Figure 5.23), although this difference was, again, slight. In general, we can see that the older group had a smaller distribution of range values compared to the younger writers.

Figure 5.26

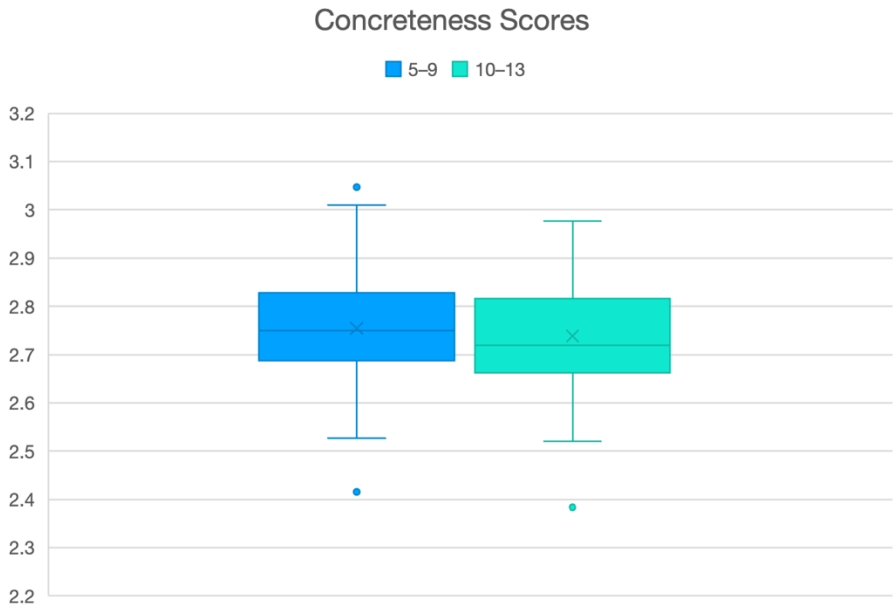
BNC Written Range Values for the Winners' Corpus



Concreteness, based on ratings collected by Brysbaert et al. (2014), slightly decreased from the younger to the older age group (mean values of 2.75 to 2.74). This meant that there was a very slight and non-significant tendency to use more abstract words with age. Research shows that abstract words tend to be acquired later in development (Ponari et al., 2018) and that children use a greater proportion of abstract words as they mature (Nippold et al., 2005; L. Sun & Nippold, 2012). Therefore, it is perhaps surprising there are not more differences in the concreteness scores between the age groups. Perhaps if the ages were able to be examined separately, we would be able to observe more fine-grained differences between students.

Figure 5.27

Concreteness Scores By Age Group In The Winners' Corpus



5.4 Discussion

Writing development is a complex phenomenon influenced by psychological and physical processes, social situations, motivations and self-perceptions of writers (Bazerman et al., 2017). Development begins in childhood and continues along a winding personal trajectory across a person's lifespan throughout ever-changing contexts (Herrington & Curtis, 2000). Writing development is singularly individual, shaped not only by educational and sociocultural opportunities, but also writers' own personal interests and abilities. The complex and multidimensional portrait of writing development means that research on its development has been fragmented. Quantitative corpus linguistic approaches to studying writing development are only one small part of the landscape on writing research. What QCL offers, however, is the ability to systematically analyse large amounts of data and evaluate patterns of language use in different contexts. This study sought

to employ QCL to examine a large corpus of children's writing collected for one specific context: fictional short stories produced for a large national writing competition.

This project had the following research questions:

1. What characterises the writing that children submit to the BBC 500 Words short-story writing competition?
2. What characterises the writing of the winners of the short-story competition? To what extent do the winners' stories differ from the other entries?
3. To what extent does the writing of the 5–9 group differ from the writing of the 10–13 group?

Several things characterised the writing submitted to the BBC 500 Words short-story writing competition (RQ1). Firstly, the themes present in these samples of writing by children echo the themes present in their reading materials, as indicated by the keyword analyses and most common parts of speech. This supports the findings of prior corpus-based research into children's literature which found core vocabulary common in child-directed texts (Dawson et al., 2021; Thompson & Sealey, 2007; Wild et al., 2013). These themes are in obvious contrast to writing produced by and for adults.

Secondly, some specific linguistic patterns were observed: children write more in the first person and they employ specific low-frequency vocabulary that does not occur to the same extent in other English corpora. The words occurring in the OCCW more than in the other corpora represent semantic areas that are particularly relevant to children, such as people, places, and nature. Their writing shows children's interest in fictional creatures, monsters, animals, and magical objects. The adjectives children use enable them to describe real and fictional worlds in vivid and imaginative ways, as things may be *magical*, *fluffy*, *scary*, or *weird*. These unique features of children's writing demonstrate the creativity of children and their varied repertoires of vocabulary items.

Furthermore, in the Sketch Engine age-related analysis, it was shown that older writers accurately employed significantly more sophisticated vocabulary, often around darker, more adult themes, going some way towards answering RQ3.

- Children use more verbs and fewer nouns in their writing compared to adult corpora;
- Children use a variety of specific low-frequency vocabulary;
- Children write about magical and fantastical things as well as places common to them, such as home and school;
- Older writers employed more sophisticated vocabulary.

Several other features characterised this corpus. The average values for the parts of speech in this corpus of children's writing were similar to the corpus of children's reading, and also aligned with figures from a corpus of children's fiction (research by Thompson and Sealey (2007), who found 15.29% nouns, 23.07% verbs, 5.29% adjectives, 7.59% adverbs, and 9.06% proper nouns and pronouns). These comparisons show that the writing children produce is similar to the writing produced for children in terms of the distribution of different parts of speech.

Where children's writing differs is when it is compared to adult-directed English. Children use more verbs and fewer nouns in their own writing than in adult fiction. This is likely due to children's focus on physical space, direct speech acts, and movement through worlds, as well as expressive language close to the self (Britton, 1972) and thus there is a reliance on core verbs to express these actions (Thompson & Sealey, 2007; Wild et al., 2013). There were only minor differences when parts of speech were compared in relation to age, with older writers using more adjectives and adverbs. Overall, there are more verbs in the OCCW than are present in the OCCR, suggesting that children write more verbs than they encounter in their reading materials; however,

in the CLIIP corpus of children's fiction, there was a slightly higher percentage of verbs (23.07% versus 22.07% in the OCCW), showing that the number of certain parts of speech depends on the corpus chosen. Perhaps the OCCR contains fewer verbs due to including non-fiction and school materials. In either case, these figures show that writing produced by children is unique in its PoS distributions, especially when compared to adult-directed sources of English. This difference when drilling down into parts of speech was also highlighted by Durrant and Brenchley (2019), who found that while overall frequency scores for texts did not show reliable age effects in their study of a corpus of children's writing (literary and non-literary texts), when they looked at different parts of speech, they found that while mean frequencies of adjectives, adverbs and verbs decreased as children aged, the mean frequencies of nouns increased as children aged. This suggested that younger children made more repeated use of low-frequency (in adult discourse) nouns, but used more high-frequency adjectives, adverbs and verbs. They summarise by emphasising that word frequency effects cannot be separated from grammar, and that lexical sophistication is inherently a multidimensional construct.

In terms of measures of lexical diversity and sophistication, this research found that as children aged (RQ3), they wrote more sophisticated and diverse texts, and also used more abstract vocabulary. This supports previous research (R. A. Berman & Nir-Sagiv, 2004; S. A. Crossley et al., 2011; Malvern et al., 2004; Olinghouse & Wilson, 2013; Uccelli et al., 2013). In general, the diversity and sophistication results reflect the literature showing that lexical diversity and sophistication increase with age. For example, regarding diversity, Olinghouse and Wilson (2013) found mean MTLD scores of 61.67 for narrative texts for their cohort of 10–11-year-olds (n = 105); the mean scores for MTLD in this research were 74.5 for the older group and 70.6 for the younger group. That these scores are higher than those calculated by Olinghouse and Wilson, even for the younger age group, bears some consideration. Perhaps the task (a prescribed story writing task from the Test of

Written Language-3, Form B versus a free-writing creative story) led to differences in scores, or perhaps it was due to text length or some other methodological consideration (such as the writers of these stories being helped by adults).

- Lexical diversity increased as children aged;
- Lexical sophistication increased as children aged, although this was more nuanced;
- Older children generally used more low-frequency words;
- Older children used more words with lower range scores;
- Concreteness decreased slightly as children aged.

Unfortunately, the winners' corpus constructed in this research was not sufficient to answer RQ2 using Sketch Engine. Thus, specific linguistic patterns separating the winners from the rest of the participants of the BBC 500 Words short-story writing competition could not be identified. Perhaps not much separates them linguistically, and the choice of winners is down to some other aspect that could be measurable if the scoring rubric was able to be examined; perhaps what drives the choice of the winners is not measurable at all, or is more down to overall perceived creativity or flair (maybe stories that make judges laugh, for example). Certainly, when the winners' corpus was being constructed, it was noted that several of the stories were extremely creative or topical, often set in far-away lands or dealing with specific social issues, such as gender and sexuality. One was written from the viewpoint of a tennis ball addressing Theresa May when she was Prime Minister, begging that she forbade using tennis balls in sports.

To summarise, children write using the words they know, drawing on a reservoir of vocabulary learned through interactions (Hoff, 2003, 2006) and through other opportunities such as reading, often with adults or teachers directing this activity (Corden, 2007; Gardner, 2004; Logan et

al., 2019). Later, their vocabularies grow through formal schooling as they learn a wider range of vocabulary, including technical and academic words. However, children also encounter words from all around them, including spoken vocabulary through conversations with peers, siblings, and adults, or passive vocabulary from television and other media. The differences found in this research between the writing produced by children and the texts produced for children show us that children's vocabularies may be sensitive to input other than reading, as the corpus of writing is unique in comparison to the corpus of reading in several key ways. These differences demonstrate the usefulness of this large, genre-controlled corpus as a source of information about the linguistic patterns in children's writing in one particular context.

5.5 Limitations and Future Work

Several methodological limitations were encountered during this research. The corpora on Sketch Engine cannot be assessed all together, i.e., although one of the benefits of this corpus is its size, the entire 440-million-word corpus cannot be analysed at once due to being separated into a separate corpus for each year of the competition on Sketch Engine. The raw data cannot be downloaded offline, and so this research could only work with Sketch Engine tools, apart from the mini winners' corpus which was created so that some offline lexical analyses could be run. Regarding this mini-corpus, it was tenuous to use this to approximate a measure of quality. The stories are not assessed systematically per se, but are evaluated by hundreds of volunteers using a pre-designed rubric assessing characterisation, plot, originality, language, and enjoyment, according to the BBC 500 Words website.

Sketch Engine is a useful piece of corpus analysis software, but is limited in its use, presenting fairly decontextualised data. The word sketches and concordance lines presented in this chapter are interesting, but it is difficult to say whether these results are valuable for shedding new

light on children's language experience in any way. Perhaps more targeted investigations would be more useful, for examining searching for specific phrases, formulaic sequences, specific vocabulary items, or grammatical structures. Doing this was not possible within the scope of this thesis, but future work could look at using this corpus to answer more specific research questions than were designed for this study.

The age-related analyses suffered from the fact that the ages are collapsed into rather broad bands (5–9 and 10–13), when there is actually a lot of vocabulary development that occurs between the ages of, for example, 4 to 9. Future research would benefit from more fine-grained analyses of age. The corpus was only able to be used via Sketch Engine, and due to privacy restrictions the raw data were unable to be provided as desired. The ages are not included on the website of the winners of each year, so separation by age was also not possible for the mini-corpus. Thus, we had to work within the parameters defined by Oxford University Press, i.e., the age categories of 4—9 and 10—13. Of note, OUP have changed their age categories for the upcoming 2024 competition, and they are now split into 5-7- and 8–11-year-olds to fit more closely to the primary school curriculum, according to the FAQs on their website, accessed September 2024.

Although one of the benefits of this research is that it captures a piece of writing produced freely by children, without constraints of time or theme, it is important to remember that these examples of writing may not be a "typical" sample for each child as they have been produced under special conditions. Children may be trying extra hard to produce creative, sophisticated vocabulary in order to write the "best" story. We do not know if the children have been helped or had their work influenced by their parents or teachers; in fact, we have to assume that an adult has been involved in the process at some point, and even one gentle and well-intentioned suggestion from an adult surrounding word choice could impact the results of this work. Spelling errors are another limitation. Although adults and parents are encouraged not to alter spelling or grammar unless the

meaning is intelligible, as per their official 2020 FAQ ("[Adults] may correct spelling where the meaning of the word might otherwise be misconstrued. However, we ask that none of the words used are changed. Grammar also has to stay the same. […] Stories are marked on their creative use of language, not on correct spelling."), it is impossible to know whether and how much an adult has assisted a child writing a 500 Words story. This echoes the research conducted with schoolteachers in Section 4 of this thesis, which found that parental involvement was a key factor in the success of children throughout the coronavirus pandemic.

The OCCW also has a number of factors that limit our interpretations of this corpus. First are the natural limitations that arise from necessary data handling procedures i.e., as all the data entered into Sketch Engine is anonymised, we do not know where the stories came from or anything about the writers apart from the county-level. This means we are not aware of several factors that could affect their vocabulary: whether they are L1 or L2 speakers of English, their socioeconomic status or parental education levels, etc. We also know nothing about what they've learned about writing, especially narrative writing, or what language they've generally been exposed to. We can assume that the majority of participants have been following the English curriculum, but how this is delivered varies by school. This is a limitation compared to studies where the subjects usually come from the same school, and so vocabulary exposure is more controlled for. However, the size of the corpus on Sketch Engine does go some way in tempering these limitations. Furthermore, vocabulary exposure cannot really be controlled for at any rate, as children will be exposed to differing amounts of language depending on the media they consume and the people they interact with.

The corpus is probably not representative of children's writing as a whole; it is, essentially, a self-sample of children who either independently chose to submit writing to this corpus or were encouraged by their teachers/parents. This competition may also naturally attract children who enjoy writing in their spare time and as such employ a more varied vocabulary. Or, inversely, they

may not enjoy the task yet were encouraged to submit by an adult in their life. Finally, the corpus is not openly accessible and access must be granted by Oxford University Press.

The OCCW contains data from the last nine years and this investigation found evidence that parts of speech frequencies as well as the most common nouns, verbs, adjectives, adverbs and proper nouns have remained relatively stable. However, this evidence is both decontextualised and preliminary. A deeper investigation into the uses of particular words, backing up the data that is produced tracking language change in the annual Oxford Language Reports, could reveal specific changes that have occurred over time and give further answers as to whether and how children's writing in this context has changed over the last nine years.

This corpus is still a rich source of information that should be more deeply explored and future research would benefit from some careful methodological considerations. If the raw data of the corpus were available offline, it could be analysed more deeply for measures of lexical sophistication and diversity using more fine-grained measures and different tools, such as the ones in this study. The winners' stories would also benefit from a more fine-grained parts of speech analysis, as the trends found in this research could be different if each part of speech was analysed separately. A final of note is that some children submit texts which cannot be classified as typical narrative prose, but rather submit stories formatted as poems, songs, letters, etc. Thus, while this is an example of lexical richness research that does control for genre to some extent, removing the stories that are not typical examples of narrative writing may have an impact on some of the results.

5.6 Conclusions

The Oxford Children's Corpus is a useful resource for studying a large sample of writing by children from all ages. This study found that children's writing overlaps with corpora of texts written for children in some areas but shows differences in others. There is less overlap for nouns and

adjectives, with some words characterising the children's writing far more than they occur in any other corpora. These words represent children's interest in the world around them and showcase their creative use of the English language; the unique vocabulary employed in their writing suggests that portions of their vocabulary are sourced outside of school and reading. In the smaller, bespoke winners' corpus, age-related differences in lexical diversity and sophistication were shown. When comparing the corpora on Sketch Engine I found that the older writers used much more complex vocabulary. Some examples of this sophisticated and diverse vocabulary were found in the older writers' corpus in Sketch Engine, such as pain, grief, and anxiety. Future work would benefit from carrying out more fine-grained analyses of this corpus, especially if the corpus was able to be analysed using programmes other than Sketch Engine itself. This is likely one of the largest corpora of original, productive, written data from children and thus offers much potential for further research into child language development.

6. Discussion and Limitations

Language defines what it means to be human—both literally and figuratively. Through words we make meaning of ourselves and the world around us. Across the globe, typically developing children *always* acquire language in a process that is one of the most remarkable feats of the human mind. But tracking the development of language is a supremely complex process. How do you begin to quantify how much language a person has? Does bigger equal better? These are the core questions at the heart of this thesis.

At this stage it is pertinent to consider the original aims of this project, conceived in 2018–19. This next paragraph is copied from the abstract of the original research project for which I applied, with the remainder of the proposal included in Appendix 3.

> Vocabulary is one of the key predictors of academic success and vocabulary deficits can limit a child's educational development. The 'word gap' has been a pervasive issue in vocabulary research since the nineties and remains an issue today. However, our understanding of these issues is limited by a lack of normative figures for vocabulary growth and development in the early years. Currently few vocabulary tests exist that can give an accurate vocabulary size for L1 English children, especially for those who are pre-literate, and so the vocabularies of younger school-age children are underreported.

It is true that the idea is pervasive. Section 2.1 demonstrated the extent to which the word gap ideology has infiltrated into UK educational policy. It is also true that lack of normative figures for vocabulary growth does limit our understanding of these issues. However, it is now apparent that the original premise of this thesis was flawed. In the following discussion I will argue that the idea of the 'word gap' or 'vocabulary gap' should be rejected as unfit for purpose. I will discuss how this gap is extremely complex to measure. I will argue that the idea of the gap is rooted in intersectional issues of poverty, classism, racism, ableism, and capitalism.

This chapter first presents a summary of this thesis and discussion of the results and their pedagogical implications for teachers and institutions. This chapter then goes on to discuss what links all these chapters together: language policy. Limitations and suggestions for future research are then summarised. It is in this way that this discussion will contribute to the literature on vocabulary development in children speaking English as an L1.

At the beginning of this thesis, I posed the question 'how can we measure the vocabularies of children in the United Kingdom?'. This question was contextualised within the English national curriculum, where vocabulary is posited as a key component of children's language learning in the UK. Each chapter of this thesis is an example of the methods which can be used by researchers to investigate children's vocabulary learning in the UK: directly testing their vocabulary, talking to teachers, and analysing their writing. This work highlights the complexity involved in quantifying children's language, and the issues and theories that must be considered before any sort of quantification can occur. For example, this thesis shows how important it is to be specific about seemingly simple concepts, such as what constitutes a word. It also shows the importance of considering research into children's vocabulary in *context*. This context is where language is made, where language is 'done', but often researchers have neglected consideration of this context, leading to sweeping claims and generalisations about language use in children.

This research is firmly contextualised within the United Kingdom, where English teaching and learning is largely controlled by the state. Thus, what connects the three chapters in this thesis is language policy. It is clear that how children are expected to 'do' language is governed by implicit and hidden agendas put forward by various groups, institutions, committees, and people. Critical discursive approaches to language require the collection and triangulation of different data sources to show how language policy does not just happen in statutory documents, but is reproduced "across multiple layers of language policy interpretation and appropriation" (Johnson, 2012, p. 119).

There are three levels to analysing language policy: 1) the micro-level (e.g., teachers, parents); 2) the meso-level (e.g., school management); and 3) the macro-level (e.g., government-produced curricula) (Cushing, 2020). These levels interact with each other to design, implement and enact policy in the UK.

It is through "overt and covert mechanisms" that those in authority (for the most part) control language spaces, as explained by Elana Shohamy in *Language Policy: Hidden agendas and new approaches*:

> the study of LP [language policy] should not be limited to formal, declared and official policies but rather to the study of the powerful mechanisms that are used in most societies nowadays to create and perpetuate "de facto" language policies and practices. The mechanisms discussed in this book include language education policies, language tests and language in the public space; it is through these policy devices that "real" policies are created. (Shohamy, 2005, p. xvi)

This thesis reflects these mechanisms: (1) Chapter 3 uses a language test to evaluate vocabulary size; (2) Chapter 4 looks first-hand at how learning changed during the pandemic, how language education policies had to change, and considers the role of teachers and parents in children's vocabulary learning; (3) Chapter 5 considers a corpus of writing produced by children across the UK written for a public short-story writing competition, one way in which ideas about children's language are brought into the public space.

Woven throughout this thesis is the underlying assumption that more language means better outcomes for students. This is shown by this idea of a large vocabulary gap separating low-performing and high-performing students, and, as the other side of this coin, the idea that having less language (however that is defined) means you are poorer or worse off in some way. Linguistic ideologies like the word gap can be harmful, especially in regard to demonising and pathologizing

minoritised and racialised groups and their language practices. How policy interacts with educational outcomes and how potentially problematic ideologies are birthed and perpetuated is evident in this research. The perpetuation of word gap discourses by respected institutions, such as Oxford University Press, and their inclusion in England's national curriculum, represents how these ideologies are made practice. In each chapter of this thesis, ideologies have been revealed pertaining to how we can, in the crudest manner, distinguish between "good" and "bad" language use by children. We will now consider further how the word gap discourse has been revealed as a hidden mechanism of language policy in the research carried out in this thesis.

Chapter 2 presented the literature review defining and summarising the idea of the word gap and presenting key findings in vocabulary testing and vocabulary research in children's writing. Both the development of literacy skills and writing itself in children were explained. Key findings from the literature included the following:

(a) The word gap was first popularised by Hart and Risley (H&R) in 1995; H&R suggested that children from lower-income households heard substantially fewer words than their higher-income counterparts, and that this was what perpetuated the cycle of poverty and led to worse outcomes for these children (Hart & Risley, 1995);

(b) Critiques (Avineri et al., 2015; Baugh, 2017; Cushing, 2023; Dudley-Marling & Lucas, 2009; Kuchirko, 2019; P. Nation, 1995; Sperry et al., 2019b; Yandell, 2017) of the influential H&R study are becoming increasingly recognised, but still its limitations are often not addressed in the literature, in official language policy documents in the UK, in publications by institutions, or by politicians;

(c) These limitations include: only considering caregiver–child interactions; the small sample size of six Black welfare families whose linguistic practices were compared

against 12 white upper-class families; the underlying raciolinguistic bias and deficit-based view (i.e., pathologizing the language practices of minorities);

(d) A near-replication of the original study could not confirm Hart and Risley's findings (Sperry et al., 2019b) and another attempted replication (Gilkerson et al., 2017) found that the gap was 4 million words (much smaller than Hart and Risley's proposed 30 million);

(e) The national curriculum for English has been influenced by word gap ideologies (conceptualised as children being 'word-rich' or 'word-poor'), suggesting that children coming to school with smaller vocabularies will be limited in their academic performance (Department for Education, 2014; Oxford University Press, 2018, 2023);

(f) Explicit descriptions of vocabulary development are important, both as a means of clarifying what a vocabulary gap might actually entail and for ensuring it is effectively targeted (Durrant & Brenchley, 2019a);

(g) There are few reports of vocabulary size in school-age children, especially in the UK; the methods used to collect vocabulary sizes vary widely, and there are several important key considerations when designing vocabulary tests, such as how to define a word and where to source words from depending on audience; the extant literature is variable, but suggests that children gain on average learn roughly 1,000 word families per year after the age of two (Brysbaert et al., 2016; Coxhead et al., 2015; D. Duff & Brydon, 2020; P. Nation & Anthony, 2016; Schmitt, 2014; Segbers & Schroeder, 2017);

(h) As children age, measures of lexical diversity and lexical sophistication in their writing increase (Durrant & Brenchley, 2019b; Johansson, 2008; Kyle et al., 2020, 2021) and they write about different things, with older writers using more abstract vocabulary, for example.

Chapter 3 presents the results of an investigation into the vocabulary sizes of 40 schoolchildren in Wales using the Picture Vocabulary Size Test (P. Nation & Anthony, 2016). The results showed that vocabulary sizes increased with age, as expected, but that there was a fair amount of variation in score, with some younger children achieving scores on par with older children. The scores in this sample contradict the larger estimations of vocabulary size growth (e.g., 1,000 word families per year) from previous research and suggest that children learn around one new word a day. The rate of learning slows down as children age, which means that the vocabulary learning load may be lower than previous literature has suggested.

However, because these scores are a decontextualised cross-section in one school, and are not linked to any demographic or socioeconomic metadata, it was not possible to draw any conclusions about what might separate the children in terms of vocabulary size. The original intention was to track development over time, which would have enabled tracking how children's scores changed over time. However, the COVID-19 pandemic meant that working with schools in person was made very difficult and so this chapter instead is a first-hand account of the difficulties of adapting research to the COVID-19 pandemic.

Chapter 4 presents the adapted programme of research that went on in schools: a set of semi-structured interviews with five teachers in the UK, collected post-COVID. These interviews provided a rich source of qualitative data. Key themes which emerged from these interviews included how teachers had to quickly adapt to changing circumstances, a general sense of anxiety and upheaval, increased workload, and the importance of relationships in a school setting. The strength and resilience shown by the teachers in these interviews reflects other research showing how important it had been during the pandemic to adapt curricula and expectations for students (Chapman et al., 2021; Orlov et al., 2020; Rehman & Fatima, 2021).

Teachers reported a strong influence of parental involvement on their students' perceived success during the pandemic. The interviewees reported that those with more involved parents performed better and had a lower perceived learning loss. These interviews demonstrated that teachers did believe in some sort of gap in the learning loss between children in the context of adjusting to life and education post-COVID. This gap was mostly observed between children who had been supported by their parents during the pandemic and those who did not receive as much help at home.

Measuring the perceived learning loss caused by the pandemic continues to be a topic of interest for policymakers, researchers, and educators. Current research suggests that the pandemic has led to a substantial learning deficit and that these deficits are particularly large among children from low socio-economic backgrounds (Betthäuser et al., 2023). It is interesting to observe how a deficit-based discourse emerged as a theme across the interviews, with a distinction made between those children who had parental support and those who did not. We can recall the observation of one respondent (R1):

> "Some parents haven't done anything at all […] we've got the children who've come on massively because the parents have been doing everything and more with them. And then we've got children who have literally done nothing. And then it's like starting from the start then."

This supports prior findings showing that parental involvement can be a key explanatory variable in educational success for children (Goodall, 2013; Goodall et al., 2021). This chapter also shows examples of barriers to parental involvement as described by prior literature (Hornby & Blackwell, 2018). For example, some parents had to continue to work during the pandemic, and were not able to be as involved with their children's education at home. Others were furloughed, spending more time at home than usual, and were able to be more involved in their children's

education, reflecting literature reporting that parental involvement increased during the pandemic (Parentkind, 2022). Some teachers commented that children who had 'difficult home lives' were more affected by learning loss in the pandemic. It is not possible to link these statements to income disparities, but these findings still have important implications, highlighting the importance of the connection between home and school. Furthermore, schools must be aware of barriers to increased parental involvement, such as uncertainties about getting involved in their children's education (due to perceived lack of knowledge or understanding) or technological barriers.

Teachers are a vehicle for ideologies, operating at the 'meso'-level of language policy (Cushing, 2020; Shohamy, 2005). While the 'word gap' was not explicitly referred to by the teachers interviewed in this study and teachers did not explicitly discuss socioeconomic situation, a similar idea emerged framed around general perceived learning loss dependent on parental involvement. We must be wary when we interpret these statements, because the teachers are only reporting observations they have made, and may not have a full and accurate understanding of the environments in the homes of their students. However, it is clear that parents as well as teachers are a key mechanism involved in language learning.

Vocabulary loss has also been specifically linked to the pandemic by educational institutions in the UK as well as the wider press. For example, Oxford University Press (2023) have recently put out a report on building vocabulary in children which, based on surveys with 503 teachers and 313 parents, claims that "Almost all of the teachers surveyed agree that pupils with more support for vocabulary development at home tend to perform better academically at school" (Oxford University Press, 2023, p. 3). This research is explicitly framed against the backdrop of the 'word gap' as an "ongoing" issue that has been "potentially heightened by the pandemic" (p. 2). The building vocabulary documents that states that

> Over half of teachers report that there has been an increase in the number of children who have fallen behind in their vocabulary knowledge compared to previous years. 95% of teachers believe that school closures and disruptions during the Covid-19 pandemic contributed to a widening vocabulary gap.

The complex relationship between the governmental, non-governmental institutions, educators, and parents affects how children learn language. In this same report, Oxford University Press state that parents desire more guidance from schools on vocabulary development and summarises some key areas of development for schools. They highlight that there is the potential to foster a reciprocal relationship between home and school to improve the vocabularies of children; this is a similar sentiment expressed by several of the teachers I interviewed, who explained that communication between home and school was a key issue during the pandemic.

This report uses the series of Oxford Language Reports (Oxford University Press, 2018, 2020) discussed in Section 2.1 as background evidence for their research. These reports are, as previously discussed, all based on the original Hart and Risley study which is the source of the word gap as a concept. These reports have been criticised for being under-researched and under-referenced, relying heavily on the original H&R study without considering its critiques (Cushing, 2023). These are further examples, then, of how educational institutions are perpetuating the word gap discourse in the UK. When these sorts of statements are put forward by influential institutions such as Oxford University Press, it can be argued that they are a mechanism “which has (re)normalized word gap ideologies in England” (Cushing, 2023, p. 318). This is because, in their regular reports on children’s language in the United Kingdom, OUP have made the word gap a key theme, legitimising it in the eyes of educators and policymakers, as well as the general public (i.e., through uncritical coverage in the UK media (Cushing, 2023, p. 318)).

The ‘vocabulary gap’ is a prime example of how under-researched ideologies can capture the minds of academics, educators, and policymakers without the required interrogation, and how they can be crystallised in policy. I believe that it is important for researchers to question the validity of claims put forward regarding the word gap, especially in the United Kingdom, as there is no evidence that such a gap exists. The majority of evidence from the UK related to the word gap has come from Oxford University Press, and is not corroborated; furthermore, the original study that the idea of the word gap arose from has several methodological and theoretical flaws, such as its sampling procedures and its lack of acknowledgment for different cultural and language practices. In effect, the word gap ideology positions privileged forms of English over others. But paradigms are shifting; times are changing. Different language and cultural experiences are not things to be ‘fixed’ and there is increasing recognition of the legitimacy of minoritised language practices (Flores & Rosa, 2015). Some have claimed that there is no reason to believe in a ‘culture of poverty’ that limits the academic and vocational success of poor people (Dudley-Marling & Lucas, 2009). To think so ignores the rich language experiences possessed by children of all cultural and linguistic groups.

These ideologies are also being perpetuated in current research. Especially notable is the lack of engagement with critiques or criticisms of the original methods. Throughout the course of this thesis, I encountered a number of papers or books which cited the original H&R study seemingly as gospel. For example, Seidenberg & Macdonald (2018) open their comprehensive paper on the role of statistical learning in language acquisition with the claim:

> Researchers now know that children’s early language experience is more variable than had previously been recognized, and that these differences have substantial effects on children’s progress in learning to read, findings that have also become widely known to the general public.

Their evidence for this claim is the Hart and Risley study, apparently supported by the "extensive literature linking language experience to socioeconomic status (Hoff, 2003)" (Seidenberg & MacDonald, 2018, p. 73). The authors acknowledge some limits in their research, recognising that vocabulary assessments may underestimate children's vocabulary as they only "index children's knowledge of the standard dialect of the language used in school but not their knowledge of another dialect or language" (p. 74). They conclude their discussion on the impact of variation with the following:

> Impoverished language experience is a source of deficits in vocabulary and other areas, but enriching that experience can potentially accelerate learning without heavy dependence on instruction. We do not yet know how to do this effectively, though researchers are exploring methods that include promoting reading to children, providing training to parents, and monitoring rates of child-directed speech, as well as prompts for child-directed conversation in public spaces such as grocery stores (Hirsch-Pasek et al., 2015; Ridge, Weisberg, Ilgaz, Hirsh-Pasek, & Golinkoff, 2015; Suskind et al., 2016). New approaches that have different or broader impacts are likely to emerge from these efforts.

It is clear, then, that ideas surrounding "impoverishment" in language are still abound. As research moves forward in this area, it is important to be mindful of perpetuating linguistic discourses which penalise marginalised and underprivileged speakers of English for not living up to the white, educated standard of English. Rather than assuming a causal link between vocabulary size and socioeconomic status, in either direction, we have to consider vocabulary knowledge and development in a wider context of intersectional issues of capitalism, poverty, classism, ableism, and racism.

Educational settings are complex and the factors which affect word learning are difficult, if not impossible, to enumerate or control. The multiplicity in education, the tangled web of

knowledge and policy which underpins a child's development, shows how important it is to work within the "mess of these multiple, inter-imbricated ways of knowing children, language and literacy in classrooms" (Burnett et al., 2020, p. 126). But children are obviously not just affected by what they learn in school; Burnett et al. argue that "engaging with the social injustices associated with language and literacy in early childhood education requires us to move beyond a perspective that starts and stays with children in classrooms" (p.118) because the moment-to-moment unfolding of events in classrooms is connected to things happening at other times and places, e.g., school development plans, learning objectives. That is to say, language policy does not just happen in government boardrooms; policies are reinforced every time a teacher or parent polices nonstandard language practices, or when the media report on slipping linguistic 'standards'. To make problems in language (in this case, a 'limited' vocabulary) the root cause of social ills is a convenient panacea (Cushing, 2023): fix the language, fix the problems. Vocabulary is perhaps the easiest panacea of them all. Words are, after all, seen as discrete and easily packaged objects, tangible things that can be captured, taught, and counted.

As research moves forward in this area, it is not enough to rely on claims made thirty years ago and which have since been linked to raciolinguistic ideologies. Researchers "must recognise that normal, average and able are socially constructed terms that must change" (Nocella II, 2017, p. 161). We must also be wary of drawing causal relationships between environment and outcomes, especially when genetic confounders are not considered (S. A. Hart et al., 2021). Challenging deficit-based thinking is key as this field moves forward and is necessary if the field is to replace old narratives with new, more helpful, ones (Burnett et al., 2020).

Chapter 5 is another example of how language policy can manifest in public space. In this chapter, an examination of an under-researched corpus of children's writing, the Oxford Children's Corpus of Writing, is presented. This corpus is made up of short stories written by children aged 5–

13 in the United Kingdom, submitted to a national competition hosted by BBC Radio 2 and Oxford University Press. The corpus is hosted on Sketch Engine. Chapter 5 shows how Sketch Engine can be used for large-scale analyses of children's writing and highlights its limitations in assessing more qualitative measures of lexical diversity or sophistication. Cumulatively, across all years, the corpus is approximately 500 million words. Only the most recent corpus available at the time of conducting this research was used (the 2020 corpus, approximately 50 million words) due to Sketch Engine's limitations. To facilitate more quantitative approaches, a smaller corpus was made using a selection of the freely available winners' stories, which are published online every year. This corpus was analysed using the Tool for the Automatic Analysis of Lexical Sophistication (Kyle et al., 2018).

The results I found support prior research (Chipere et al., 2001; Durrant, 2022b; Durrant & Brenchley, 2019b; Johansson, 2008; Kyle et al., 2020; Olinghouse & Wilson, 2013) showing that measures of lexical diversity and sophistication in children's writing samples increase as children get older. The results showed that children employ a variety of vocabulary in their narrative writing, using more low-frequency nouns (such as *dragon* or *fairy*) in comparison to adult corpora, and writing about topics close to home, such as their family. Furthermore, children's writing was characterised by an emphasis on action, physical space, and the natural world, similar to the themes found in children's literature compared to adult's literature (Thompson & Sealey, 2007; Wild et al., 2013). Children wrote with a strong narrative voice, echoing work on children's literature (McCallum & Stephens, 2011; Stephens, 2005; Wall & Crevecoeur, 2016), and positioned themselves firmly within relevant 'institutions', e.g., family, friendship, gender, home, race, and religion (Knowles & Malmkjær, 2002). Through writing about these institutions, the child authors are able to 'locate' the reader while representing their own world view.

Older writers used more complex vocabulary and the shift to more 'adult' themes is clearly marked when you compare the older writers to the younger writers. Research in the same Oxford

Children's corpus has shown that positive sentiments in children's writing decrease with age (Dong et al., 2024). This was reflected in the findings of the research conducted in this thesis, with older writers using words such as *suicide*, *anxiety*, *depression*, *grief* and *murder* where younger writers did not. Further research could explore how and when these darker themes emerge in children's writing.

Competitions like the BBC 500 Words are effective ways of encouraging children to be creative in their language use, and enable the collection of a large amount of linguistic data. The Oxford Children's Corpus, where the entries to this competition are collated, is an extremely rich source of data which hitherto has been relatively unexplored by researchers. One study used it to show a tendency for boys to prefer writing about boys, whereas girls were more balanced in their choices to write about male or female characters, mirroring adult authors (Hsiao et al., 2021). To my understanding, this study used metadata from the corpus pulled from Sketch Engine. A more recent study has used the writing portion of the Oxford Children's Corpus to show that positive sentiments in children's writing decrease with age (Dong et al., 2024).

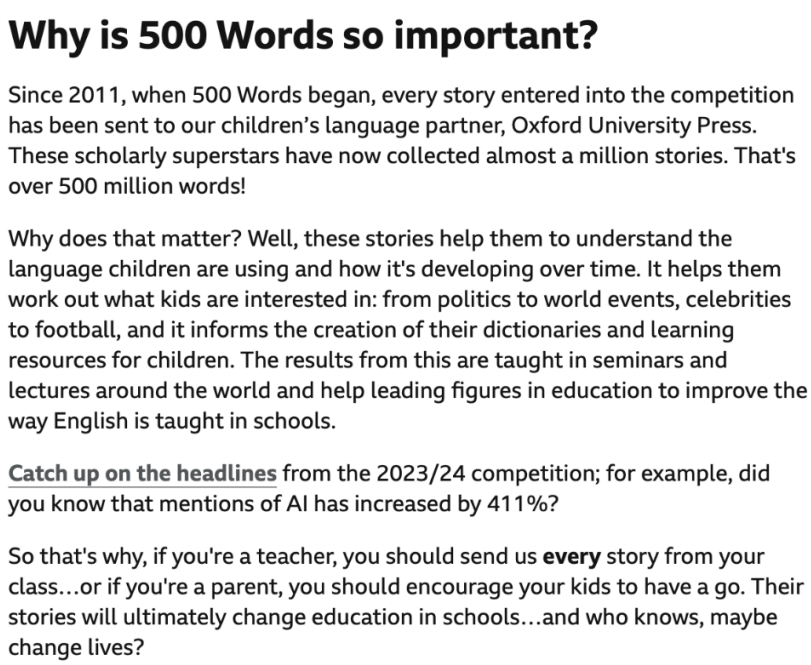
Oxford University Press evaluate and report on the words that children use in their corpus every year, providing 'Word of the Year' reports based on the corpus made up of entries to their competition. In their most recent analysis of the stories submitted to the 2023 competition, OUP include sections on children's creative use of metaphor, unusual story openers, and invented words (Oxford University Press, 2024). Their reports highlight how children are in touch with current affairs, with the most recent report, for example, highlighting a significant increase in mentions of conflict (p. 3) due to current affairs (i.e., the Ukraine war and the Israel/Palestine conflict). It is encouraging to read these reports and see how children are engaged with creative writing in this way. These reports highlight the variety and creativity in children's vocabularies and my research has also shown how children become more adept at using more sophisticated language as they age.

However, it is important to note several things about the reports that OUP publish every year on the language used in the BBC 500 Words competition. Firstly, OUP compare the most recent corpus with the previously available one, reporting year-on-year differences. So, while they report significant increases, e.g., an increase of +1914% in the word *Gaza* from 2020 to 2023, this could represent an increase of 1 occurrence in the 2020 corpus to 20 occurrences in the 2023 (with the corpus in total having around 100,000 stories). The exact occurrences are not reported. Secondly, it is important to highlight that this is counting the number a times a word occurred (in this example, *Gaza*), and these words could all have come from the same story. Thirdly, while findings like this are interesting, they are not surprising; to give another example, they report that "use of the word AI in stories from 2023 is five times as frequent as in 2020 (+411%)" (p. 7). This makes sense when we remember that AI is a relatively recent phenomenon.

These reports are examples of how ideas about language are brought into the public space and perpetuated by institutions in the UK. In this case, the word gap ideology emerges in (a) OUP's annual reports, which position the corpus as a primary source of data on how children use language, despite a lack of empirical research into the corpus; and (b) the language used when discussing the 500 Words Corpus itself. For example, the frequently asked questions section of the 500 Words website contains an explanation that the reason the competition is so important is because the corpus constructed from the entries to the competition helps researchers understand what language children are using and how it is developing over time, to "help leading figures in education improve the way English is taught in schools […] and who knows, maybe change lives?" (Figure 6.1). However, how the corpus is being used in this manner is not clear, beyond a few empirical publications and the reports published by Oxford University Press themselves.

Figure 6.1

Excerpt From BBC 500 Words Website: https://www.bbc.co.uk/teach/500-words/articles/zctk7v4#z3qk239 (accessed 11 June 2024).



Why is 500 Words so important?

Since 2011, when 500 Words began, every story entered into the competition has been sent to our children's language partner, Oxford University Press. These scholarly superstars have now collected almost a million stories. That's over 500 million words!

Why does that matter? Well, these stories help them to understand the language children are using and how it's developing over time. It helps them work out what kids are interested in: from politics to world events, celebrities to football, and it informs the creation of their dictionaries and learning resources for children. The results from this are taught in seminars and lectures around the world and help leading figures in education to improve the way English is taught in schools.

Catch up on the headlines from the 2023/24 competition; for example, did you know that mentions of AI has increased by 411%?

So that's why, if you're a teacher, you should send us **every** story from your class…or if you're a parent, you should encourage your kids to have a go. Their stories will ultimately change education in schools…and who knows, maybe change lives?

Positioning the corpus as an extremely valuable source of information which has the potential to influence policy while simultaneously maintaining significant barriers to its use echoes prior criticism of how Oxford University Press positions itself in the language policy landscape of the UK. For example, Cushing has suggested that the design and marketing of resources to schools with the aim to 'close the gap' is evidence that OUP are "working to commodify word gap ideologies and convert them into economic profit" (Cushing, 2023, p. 318). Commodification of deficit-based ideologies in general is common: Burnett et al. (2020) highlight how interventions to close attainment gaps reflect a level of high-stakes accountability and marketisation that has long "plagued" the English educational system (p. 123). This commodification seems inevitable, but my hope is that some of what I have presented in this thesis makes us question whether institutions in

the United Kingdom are drawing attention to problems purely so they can then sell the solutions to them.

So far, the corpus has been growing for over a decade, yet there are no publications from anyone not significantly connected to Oxford University Press. Lowering barriers to the corpus is especially important if it continues to be linked to influential educational policy decisions. Dong's recent study (2024) used one year of the corpus for analyses of .txt files, and this seems to have been carried out offline, indicating that perhaps the issues I encountered during this research in terms of accessing the corpus offline are being overcome. It is heartening to see some signs of moving this corpus in a more open science direction, with both the above authors providing open materials and clear description of methods. What I was able to do in this research, limited by Sketch Engine's functions, has only scratched the surface, and there is plenty that could be done with this corpus, from more detailed analyses of lexical diversity, density, and sophistication, to investigations of children's use of formulaic sequences (a relatively underexamined phenomenon). With more time and resources, even further investigation only using Sketch Engine could yield more findings, perhaps through a more fine-grained analysis of differences in word usage by age. Through comparison with books, media, or educational materials, it could be possible to identify when and how children begin using particular words.

Of course, in this era of artificial intelligence, there are also arguments to be made for protecting a valuable linguistic resource, and I cannot make comments on the legal or other aspects of releasing this corpus more widely. The corpus is certainly useful for educators and those who develop educational materials, such as Oxford University Press, as it allows them to keep their dictionaries and other materials up to date with relevant information for teachers and students (Armstrong & Banerji, 2021). Unfortunately, the restrictions placed around this thesis limit its potential applications. In my research I encountered barriers to downloading and accessing the

corpus offline, despite being given full access to the corpus on Sketch Engine. This meant that I had to work around significant limitations, and was in the end limited to using Sketch Engine's lexical analysis features rather than being able to explore the corpus in a bespoke manner. It is my hope that the barriers to this corpus will be lowered in the future, so that it can be used by more researchers as a source of naturally occurring language data for children in the UK.

Further recommendations from this work are made complicated by the layers of language policing which surround children's English development in the United Kingdom. This policing comes from the people who write the English curriculum and those who uphold it, including teachers in their classrooms, and Ofsted, the English school inspectorate. This inspectorate is predominantly made up of white, privileged inspectors who can be said to be an arm of government surveillance (Cushing & Snell, 2023). With their 'listening ears', Ofsted inspectors monitor the language that both teachers and students are using in schools, penalising those who step outside of the 'acceptable' language box. But children's language is diverse, and should be celebrated for its diversity; teachers can be encouraged by the research in this thesis which shows that a wide variability in vocabulary knowledge is normal in growing children.

Researchers must continue to work together to decolonise our field, pushing back against deficit-based narratives and ensuring that we do not uncritically accept ideas like the vocabulary gap merely because they seem easy to fix or because they emotionally resonate with us. We must be wary of using results connecting socioeconomic status with vocabulary size, or vocabulary size with academic success, without considering wider intersectional issues. Of course parents struggling to make ends meet might talk less to their children; of course children who are not adequately fed might not be so interested in improving their vocabularies. This does not mean that poorer parents are bringing their children up in linguistically 'impoverished' households.

Policymakers should be equally mindful. The complexity of literacy, and its interactions with social and cultural values, as well as the individual differences that writers from different societies bring, mean there is not a one-size-fits-all writing curriculum, because writing develops and changes throughout childhood, adolescence, and adulthood (Bazerman et al., 2018). Future curricula must address the diversity of students' needs, abilities, interests, and strengths. Throughout this thesis, England's National Curriculum has been used as an example of how literacy practices are implemented in the United Kingdom. Education in the United Kingdom is, however, a devolved issue. This means that Wales, Northern Ireland, and Scotland have separate educational policies and systems under separate governments. In April 2021, the Welsh government passed the Curriculum and Assessment (Wales) Act (2021), establishing a new framework for the curriculum for pupils in Wales. This new Curriculum for Wales, which is being gradually phased in at the time of writing, represents a significant departure from the much more prescriptive guidance of the national curriculum in England. Control of curriculum and assessment design and implementation is given back to schools, with head teachers responsible for designing and implementing the curriculum in their school. The new Welsh Curriculum promotes language and literacy as a tool that can help students become informed, global citizens, and that acknowledges that children progress at different rates (Welsh Government, 2020). Literacy is enshrined as a "mandatory cross-curricular skill" in the Literacy and Numeracy Framework. Multilingualism is emphasised alongside the importance of valuing all languages and cultures.

It is not yet clear what impact this new curriculum will have on literacy education in Wales, but its implementation heralds a potential new curriculum for England, too, to be implemented by the new Labour government. If so, it is imperative that the curriculum designers are made aware of long-standing raciolinguistic ideologies so they can effectively decolonise the curriculum, exposing

ways in which white privilege exerts its invisible power (Arday et al., 2021; Begum & Saini, 2019; Moncreieffe et al., 2020).

7. Conclusions

Vocabulary knowledge is a multifaceted and complex aspect of a child's linguistic development, and quantifying it is difficult; vocabulary growth is dynamic, reliant on a myriad of internal and external factors. Through a comprehensive examination of empirical studies framed against the backdrop of language educational policy and ideologies in the United Kingdom, I have presented several ways in which the vocabulary knowledge of children can be measured. I have shown that children in the United Kingdom demonstrate remarkable variety in their vocabulary use, as well as resilience in the face of adversity (i.e., the COVID-19 pandemic).

Though this thesis did not answer the original question it set out to investigate, it makes several contributions to the field of applied linguistics. Distinct methodologies for investigating vocabulary in L1 English-speaking children have been employed in this research, and their strengths and weaknesses discussed. Quantitative methods, such as vocabulary testing and quantitative corpus linguistic approaches, can provide information on the words children know and use. Here, I was able to highlight the diversity of vocabulary known and employed by children in the UK. Qualitative approaches, such as interviews, provide the context for how children learn words. My interviews captured a unique context: how teachers in the UK adapted to the impact of COVID-19 pandemic and its subsequent school closures. These interviews exemplified the commitment and resilience that teachers show, and also highlighted how important the connection between home and school is. Furthermore, the interviewees all discussed how children with more external support at home seemed to fare better during the pandemic.

The overarching contribution of this thesis is its investigation into the word gap. Building on a body of literature which is increasingly critical of the vocabulary gap, the reflections on the methodological implications of investigating the word gap in this thesis encourage a more nuanced view of the concept. My intention with this work is not to denounce prior research or interventions,

but to ask laypeople, policymakers, and researchers to examine their preconceived notions regarding the vocabulary gap. In trying to research the gap, it was made apparent just how complex it is to make claims about children's language development, and the potential impact such claims from researchers can have on educational policymaking across the globe. Examining naturalistic linguistic environments is methodologically complicated by issues of race, class, culture, and economics. As work in this area moves forward, it is important for researchers in early language acquisition to be cognizant of the critiques of the word gap literature to avoid perpetuating deficit-based perspectives.

It is my view that the field of applied linguistics needs to ensure that programmes are built on sound developmental principles which also respect the linguistic practices of different communities (Kuchirko, 2019, p. 356). We can hold ourselves to higher standards to justify interventions; we must be more careful about why we are designing these interventions and how we are justifying them. I hope that this work can bring awareness to this issue and encourage others to question the validity of some longstanding beliefs in linguistics. As research in this field continues to evolve, it is essential to consider the diverse contexts in which children learn, use, and 'do' language to ensure that we set all children up to acquire the vocabulary that they require to meet their needs.

8. References

Abell, J., & Myers, G. (2008). Analyzing Research Interviews. In R. Wodak & M. Krzyżanowski (Eds.), *Qualitative Discourse Analysis in the Social Sciences*. Palgrave Macmillan.

Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon*. John Wiley & Sons.

Alexiou, T., & Milton, J. (2019). Pic-Lex: A new tool of measuring receptive vocabulary for very young learners. In W. Zoghbor & T. Alexiou (Eds.), *Advancing ELT Education*. Zayed University Publications.

Allen, R., Jerrim, J., & Sims, S. (2020). *How did the early stages of the COVID-19 pandemic affect teacher wellbeing? Centre for Education Policy and Equalising Opportunities (CEPEO)*.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. *Comprehension and Teaching: Research Reviews*, 77–117.

Anderson, S., Medrich, E., & Fowler, D. (2007). Which achievement gap? *Phi Delta Kappan*, *88*(07), 547–550.

Andrew, Alison., Cattan, Sarah., Costa-Dias, Monica., Farquharson, Christine., Kraftman, Lucy., Krutikova, Sonya., Phimister, Angus., Sevilla, Almudena., & Institute for Fiscal Studies (UK). (2020). *Learning during the lockdown : Real-time data on children's experiences during home learning.*

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, i–186.

Applebee, A. (2000). Alternative models of writing development. *Perspectives on Writing: Research, Theory, and Practice*, 90–110.

Arday, J., Zoe Belluigi, D., & Thomas, D. (2021). Attempting to break the chain: reimaging inclusive pedagogy and decolonising the curriculum within the academy. *Educational Philosophy and Theory*, *53*(3), 298–313. https://doi.org/10.1080/00131857.2020.1773257

Armstrong, S., & Banerji, N. (2021). *Oxford Children's Word of the Year & Oxford Children's Language Report 2021*.

Arnold, C. (1999). *Child Development and Learning 2-5 Years: Georgia's Story*. Sage.

Attainment gap in primary schools wider since Covid - BBC News. (2021, December 14). *BBC News*. https://www.bbc.co.uk/news/uk-scotland-59650826

Avineri, N., Johnson, E., Brice-Heath, S., McCarty, T., Ochs, E., Kremer-Sadlik, T., Blum, S., Zentella, A. C., Rosa, J., & Flores, N. (2015). Invited forum: Bridging the "language gap." *Journal of Linguistic Anthropology*, *25*(1), 66–86.

BAAL. (2021). *Recommendations on Good Practice in Applied Linguistics. 4th Edition.*

Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.

Bachman, L., & Palmer, A. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests* (Vol. 1). Oxford University Press.

Baker, C., Hutton, G., Christie, L., & Wright, S. (2020, December 17). *COVID-19 and the digital divide - POST*. POST. https://post.parliament.uk/covid-19-and-the-digital-divide/

Baker, P. (2010). *Corpus Methods in Linguistics* (L. Litosseliti, Ed.). Continuum International.

Banerji, N., Gupta, V., Kilgarriff, A., & Tugwell, D. (2013). Oxford Children's Corpus: a corpus of children's writing, reading, and education. *Corpus Linguistics 2013*, 315.

Barbour, M. K., Labonte, R., Kelly, K., Hodges, C., Moore, S., Lockee, B., Trust, T., Bond, A., & Hill, P. (2020). *Understanding Pandemic Pedagogy: Differences Between Emergency Remote, Remote, and Online Teaching A special report of the State of the Nation: K-12 E-Learning in Canada project*. https://www.flickr.com/photos/51035597898@N01/544654854/

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279.

Baugh, J. (2017). Meaning-Less Differences: Exposing Fallacies and Flaws in "The Word Gap" Hypothesis That Conceal a Dangerous "Language Trap" for Low-Income American Families and

Their Children. *International Multilingual Research Journal*, *11*(1), 39–51. https://doi.org/10.1080/19313152.2016.1258189

Bazerman, C., Applebee, A. N., Berninger, V., Brandt, D., Graham, S., Matsuda, P. K., Murphy, S., Rowe, D. W., & Schleppegrell, M. (2017). *Taking the long view on writing development*.

Bazerman, C., Applebee, A. N., Berninger, V. W., Brandt, D., Graham, S., Jeffery, J. V, Matsuda, P. K., Murphy, S., Rowe, D. W., & Schleppegrell, M. (2018). *The lifespan development of writing*. National Council of Teachers of English Urbana, IL.

Bazerman, C., Krut, R., Lunsford, K., McLeod, S., Null, S., Rogers, P., & Stansell, A. (2009). *Traditions of writing research*. Routledge.

Bearman, M. (2019). Focus on Methodology: Eliciting rich data: A practical approach to writing semi-structured interview schedules. *FOCUS ON HEALTH PROFESSIONAL EDUCATION: A MULTI-PROFESSIONAL JOURNAL*, *20*(3).

Beglar, D. (2010). A Rasch-based validation of the vocabulary size test. *Language Testing*, *27*(1), 101–118. https://doi.org/10.1177/0265532209340194

Beglar, D., & Nation, P. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13.

Begum, N., & Saini, R. (2019). Decolonising the Curriculum. *Political Studies Review*, *17*(2), 196–201. https://doi.org/10.1177/1478929918808459

Behrens, H. (2008). *Corpora in language acquisition research: History, methods, perspectives* (Vol. 6). John Benjamins Publishing.

Berman, R. A. (1995). Narrative competence and storytelling performance: How children tell stories in different contexts. *Journal of Narrative and Life History*, *5*(4), 285–313.

Berman, R. A., & Nir-Sagiv, B. (2004). Linguistic indicators of inter-genre differentiation in later language development. *Journal of Child Language*, *31*(2), 339–380. https://doi.org/10.1017/S0305000904006038

Berman, R. A., & Slobin, D. I. (2013). *Relating events in narrative: A crosslinguistic developmental study*. Psychology Press.

Berman, R., & Nir, B. (2010). The lexicon in writing–speech-differentiation. *Written Language & Literacy*, *13*(2), 183–205.

Betthäuser, B. A., Bach-Mortensen, A. M., & Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour*, *7*(3), 375–385. https://doi.org/10.1038/s41562-022-01506-4

Biber, D., & Conrad, S. (2019). *Register, genre, and style*. Cambridge University Press.

Biemiller, A. (2005). *Size and sequence in vocabulary development: Implications for choosing words for primary grade vocabulary instruction*.

Biemiller, A. (2011). Vocabulary Development and Instruction: A Prerequisite for School Learning Vocabulary Development and Implications For Academic Success. In S. Neuman & D. Dickinson (Eds.), *Handbook of Early Literacy Research* (Vol. 2). Guilford Press.

Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, *37*(6), 1461–1476.

Boers, F., Eyckmans, J., & Stengers, H. (2007). Presenting figurative idioms with a touch of etymology: More than mere mnemonics? *Language Teaching Research*, *11*(1), 43–62.

Botting, N. (2002a). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, *18*(1), 1–21. https://doi.org/10.1191/0265659002ct224oa

Botting, N. (2002b). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, *18*(1), 1–21.

Bowles, R. P., & Salthouse, T. A. (2008). Vocabulary test format and differential relations to age. *Psychol Aging*, *23*(2), 366–376.

Bracken, B. A., & Murray, A. M. (1984). Stability and predictive validity of the PPVT-R over an eleven month interval. *Educational & Psychological Research*.

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. sage.

Braun, V., & Clarke, V. (2019). *Reflecting on reflexive thematic analysis*.

Britton, J. (1972). Writing to Learn and Learning to Write. In *The Humanity of English: 1972 Distinguished Lectures*.

Brouwers, A., Tomic, W., & Boluijt, H. (2011). Job demands, job control, social support and self-efficacy beliefs. In *Europe's Journal of Psychology* (Vol. 7, Issue 1). www.ejop.org

Brown, D., Stewart, J., Stoeckel, T., & Mclean, S. (2021). THE COMING PARADIGM SHIFT IN THE USE OF LEXICAL UNITS. In *Studies in Second Language Acquisition* (Vol. 43, Issue 5, pp. 950–953). Cambridge University Press. https://doi.org/10.1017/S0272263121000668

Brown, D., Stoeckel, T., Mclean, S., & Stewart, J. (2020). The Most Appropriate Lexical Unit for L2 Vocabulary Research and Pedagogy: A Brief Review of the Evidence. *Applied Linguistics*, *43*(3), 596–602. https://doi.org/10.1093/applin/amaa061

Brown, P., & Gaskins, S. (2014). Language acquisition and language socialization. In N. J. Enfield, P. Kockelman, & J. Sidnell (Eds.), *The Cambridge Handbook of Linguistic Anthropology* (pp. 187–226). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9781139342872.010

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, *7*(JUL), 1–11. https://doi.org/10.3389/fpsyg.2016.01116

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Burnett, C., Merchant, G., & Neumann, M. M. (2020). Closing the gap? Overcoming limitations in sociomaterial accounts of early literacy. *Journal of Early Childhood Literacy*, *20*(1), 111–133. https://doi.org/10.1177/1468798419896067

Cabral-Gouveia, C., Menezes, I., & Neves, T. (2023). Educational strategies to reduce the achievement gap: a systematic review. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1155741

Carrillo, C., & Flores, M. A. (2020). COVID-19 and teacher education: a literature review of online teaching and learning practices. *European Journal of Teacher Education*, *43*(4), 466–487. https://doi.org/10.1080/02619768.2020.1821184

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a Papuan community. *Journal of Child Language*, *48*(4), 792–814. https://doi.org/10.1017/S0305000920000549

Chamberlain, L., Lacina, J., Bintz, W. P., Jimerson, J. B., Payne, K., & Zingale, R. (2020). Literacy in Lockdown: Learning and Teaching During COVID-19 School Closures. *Reading Teacher*, *74*(3), 243–253. https://doi.org/10.1002/trtr.1961

Champion, T. B., Hyter, Y. D., McCabe, A., & Bland-Stewart, L. M. (2003). "A Matter of Vocabulary" Performances of Low-Income African American Head Start Children on the Peabody Picture Vocabulary Test—III. *Communication Disorders Quarterly*, *24*(3), 121–127.

Chapelle, C. A. (2012). Validity argument for language assessment: The framework is simple…. *Language Testing*, *29*(1), 19–27.

Chapman, S., Aled Evans, R., Horder, S., Jones, C., Lloyd-Williams, S., Mahoney, N., Meace, S., & Pullen, M. (2021). *Experiences of Blended and Distance Learning During the COVID-19 Pandemic in Wales: School and Stakeholder Evidence*.

Cheng, W. (2011). *Exploring corpus linguistics: Language in action*. Routledge.

Children's Commissioner. (2020). *Stress among children in England during the coronavirus lockdown*. https://www.childrenscommissioner.gov.uk/2020/04/03/angry-fed-up-isolated-coronavirus-and-childrens-mental-health/

Chipere, N., Malvern, D., Richards, B., & Duran, P. (2001). Using a corpus of school children's writing to investigate the development of vocabulary diversity. *Technical Papers. Volume 13. Special Issue. Proceedings of the Corpus Linguistics 2001 Conference*, 126–133.

Choi, K. R., Heilemann, M. S. V., Fauer, A., & Mead, M. (2020). A Second Pandemic: Mental Health Spillover From the Novel Coronavirus (COVID-19). *Journal of the American Psychiatric Nurses Association*, *26*(4), 340–343. https://doi.org/10.1177/1078390320919803

Clark, C. (2018). *Children and young people's writing in 2017/2018 Findings from our Annual Literacy Survey*.

Clark, C., Picton, I., & Lant, F. (2020). *"More time on my hands": Children and young people's writing during the COVID-19 lockdown in 2020*.

Clark, U. (2001). Nationalising a Curriculum for English. In *War Words: Language, History, and the Disciplining of English* (p. 309). Elsevier.

Coleman, V. (2021). *Digital divide in UK education during COVID-19 pandemic: Literature review Research Report*. http://www.cambridgeassessment.org.uk

Constantinou, F. (2019). The construct of language competence over time: using high-stakes tests to gain insight into the history of L1 education in England. *Language and Education*, *33*(6), 491–505. https://doi.org/10.1080/09500782.2019.1597106

Corbin, J., & Strauss, A. (2008). Strategies for qualitative data analysis. *Basics of Qualitative Research. Techniques and Procedures for Developing Grounded Theory*, *3*(10.4135), 9781452230153.

Corden, R. (2007). Developing reading-writing connections: the impact of explicit instruction of literary devices on the quality of children's narrative writing. *Journal of Research in Childhood Education*, *21*(3).

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238.

Coxhead, A., Nation, P., & Sim, D. (2015). Measuring the vocabulary size of native speakers of English in New Zealand secondary schools. *New Zealand Journal of Educational Studies*, *50*(1), 121–135.

Crossley, S. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, *11*(3), 415–443. https://doi.org/10.17239/JOWR-2020.11.03.01

Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, *18*(2), 119–135. https://doi.org/10.1016/j.jslw.2009.02.002

Crossley, S. A., Weston, J. L., McLain Sullivan, S. T., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication*, *28*(3), 282–311. https://doi.org/10.1177/0741088311410188

Curry, D. (2023, January 9). *Microsoft Teams Revenue and Usage Statistics (2023)*. Business of Apps. https://www.businessofapps.com/data/microsoft-teams-statistics/

Cushing, I. (2020). The policy and policing of language in schools. *Language in Society*, *49*(3), 425–450. https://doi.org/10.1017/S0047404519000848

Cushing, I. (2023). Word rich or word poor? Deficit discourses, raciolinguistic ideologies and the resurgence of the 'word gap' in England's education policy. *Critical Inquiry in Language Studies*, *20*(4), 305–331. https://doi.org/10.1080/15427587.2022.2102014

Cushing, I., & Snell, J. (2023). The (white) ears of Ofsted: A raciolinguistic perspective on the listening practices of the schools inspectorate. *Language in Society*, *52*(3), 363–386. https://doi.org/10.1017/S0047404522000094

Daller, M. H., & Phelan, D. (2013). Predicting international student study success. *Applied Linguistics Review*, *4*(1), 173–193. https://doi.org/10.1515/applirev-2013-0008

Davies, M. (2018). Corpus-based Studies of Lexical and Semantic Variation: The Importance of Both Corpus Size and Corpus Design. In C. Suhr, T. Nevalainen, & I. Taavitsainen (Eds.), *From Data to Evidence in English Language Research*. Brill. http://ebookcentral.proquest.com/lib/swansea-ebooks/detail.action?docID=5634141.

Dawson, N., Hsiao, Y., Wei, A., Tan, M., & Banerji, N. (2021). Features of lexical richness in children's books: Comparisons with child-directed speech. *Language Development Research*. https://doi.org/10.34842/5we1-yk94

Denscombe, M. (2017). *EBOOK: The good research guide: For small-scale social research projects*. McGraw-Hill Education (UK).

Denzin, N. K., & Lincoln, Y. S. (2011). *The Sage handbook of qualitative research*. sage.

Department for Education. (2014, July 16). *Statutory guidance: National curriculum in England: English programmes of study*. https://www.gov.uk/government/publications/national-curriculum-in-england-english-programmes-of-study/national-curriculum-in-england-english-programmes-of-study

Department for Education. (2022, July 15). *Research review series: English*. https://www.gov.uk/government/publications/curriculum-research-review-series-english/curriculum-research-review-series-english

Derrah, R., & Rowe, D. E. (2015). Validating the Japanese bilingual version of the Vocabulary Size Test. *International Journal of Languages, Literature and Linguistics*, *1*(2), 131–135.

Diller, K. C. (1978). *The language teaching controversy*. Newbury House Publishers.

Dong, Y., Hsiao, Y., Dawson, N., Banerji, N., & Nation, K. (2024). The Emotional Content of Children's Writing: A Data-Driven Approach. *Cognitive Science*, *48*(3). https://doi.org/10.1111/cogs.13423

Donnelly, R., & Patrinos, H. A. (2022). Learning loss during Covid-19: An early systematic review. In *Prospects* (Vol. 51, Issue 4, pp. 601–609). Springer Science and Business Media B.V. https://doi.org/10.1007/s11125-021-09582-6

Dörnyei, Z. (2007). *Research methods in applied linguistics*.

Dudley-Marling, C., & Lucas, K. (2009). Pathologizing the Language and Culture of Poor Children. *Language Arts*, *86*(5), 362–370. https://doi.org/10.58680/la20097099

Duff, D., & Brydon, M. (2020). Estimates of individual differences in vocabulary size in English: how many words are needed to 'close the vocabulary gap'? *Journal of Research in Reading*, *43*(4), 454–481. https://doi.org/10.1111/1467-9817.12322

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *56*(8), 848–856. https://doi.org/10.1111/jcpp.12378

Dunn, L., & Dunn, D. (2007). *Peabody Picture Vocabulary Test, Fourth Edition* (Pearson, Ed.). https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Academic-Learning/Brief/Peabody-Picture-Vocabulary-Test-%7C-Fourth-Edition/p/100000501.html

Dunst, C. J., & Gorman, E. (2009). *Development of Infant and Toddler Mark Making and Scribbling*. *2*(2).

Durrant, P. (2022a). *Corpus Linguistics for Writing Development: A Guide for Research*. Taylor & Francis.

Durrant, P. (2022b). Studying children's writing development with a corpus. *Applied Corpus Linguistics*, *2*(3), 100026. https://doi.org/10.1016/j.acorp.2022.100026

Durrant, P., & Brenchley, M. (2019a). Corpus research on the development of children's writing in L1 English. In A. Abel, A. Glaznieks, V. V. Lyding, & L. Nicolas (Eds.), *Widening the Scope of Learner Corpus Research. Selected Papers from the Fourth Learner Corpus Research Conference*. Presses Universitaires de Lovain.

Durrant, P., & Brenchley, M. (2019b). Development of vocabulary sophistication across genres in English children's writing. *Reading and Writing*, *32*(8), 1927–1953. https://doi.org/10.1007/s11145-018-9932-8

Durrant, P., & Brenchley, M. (2021). The development of academic collocations in children's writing. In *Vocabulary Theory, Patterning and Teaching*. Multilingual Matters.

Durrant, P., Brenchley, M., & McCallum, L. (2021). *Understanding Development and Proficiency in Writing: Quantitative Corpus Linguistic Approaches*. Cambridge University Press.

Education Policy Institute. (2020). *Preventing the disadvantage gap from increasing during and after the Covid-19 pandemic*.

Elgort, I. (2013). Effects of L1 definitions and cognate status of test items on the Vocabulary Size Test. *Language Testing*, *30*(2), 253–272.

Elliot Major, L., Eyles, A., & Machin, S. (2021). *Unequal Learning and Labour Market Losses in the Crisis: Consequences for Social Mobility Centre for Education Policy and Equalising Opportunities (CEPEO)*.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143–188.

Epstein, J. L. (2018). School, family, and community partnerships in teachers' professional work. *Journal of Education for Teaching*, *44*(3), 397–406. https://doi.org/10.1080/02607476.2018.1465669

Fenson, L. (2007). *MacArthur-Bates communicative development inventories*.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248. https://doi.org/10.1111/desc.12019

Flores, N., & Rosa, J. (2015). Undoing Appropriateness: Raciolinguistic Ideologies and Language Diversity in Education. In *Harvard Educational Review* (Vol. 85, Issue 2). http://meridian.allenpress.com

Ford Smith, V. (2021). Adult. In P. Nel, L. Paul, & N. Christensen (Eds.), *Keywords for Children's Literature* (Second).

Ford, T., John, A., & Gunnell, D. (2021). Mental health of children and young people during pandemic. *The BMJ*, *372*. https://doi.org/10.1136/bmj.n614

Foster, J., & Mackie, C. (2013). *Lexical analysis of the Dr. Seuss Corpus*. Concordia Working Papers in Applied Linguistics.

Fox, C. (2003). Playing the storyteller: Some principles for learning literacy in the early years of schooling. *Handbook of Early Childhood Literacy*, 189–198.

Fox, G. (2004). *Children's Literature, Drama and Oral Storytelling*.

Freedman, A., & Pringle, I. (1980). Writing in the college years: Some indices of growth. *College Composition and Communication*, *31*(3), 311–324.

Fulcher, G. (2009). Test use and political philosophy. *Annual Review of Applied Linguistics*, *29*, 3–20.

Fulcher, G. (2013). *Practical language testing*. Routledge.

Gablasova, D., & Brezina, V. (2021). Words that matter in L2 research and pedagogy: A corpus-linguistics perspective. *Studies in Second Language Acquisition*, *43*, 958–961.

Gardner, D. (2004). Vocabulary Input through Extensive Reading: A Comparison of Words Found in Children's Narrative and Expository Reading Materials. In *Nagy and Herman*.

Gee, J. P. (1987). Issue 1 Article 2 9-1-1987 Part of the Scholarship of Teaching and Learning Commons Recommended Citation Recommended Citation Gee. In *Journal of Teaching and Learning* (Vol. 2). https://commons.und.edu/tl-journal

Gibb, N. (2015). How E. D. Hirsch Came to Shape UK Government Policy. In J. Simons & N. Porter (Eds.), *Knowledge and the Curriculum A collection of essays to accompany E. D. Hirsch's lecture at Policy Exchange* (pp. 12–20).

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Oller, D. K., Hansen, J. H. L., & Paul, T. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology*, *26*(2), 248–265. https://doi.org/10.1044/2016_AJSLP-15-0169

Girlguiding. (2020). *Girlguiding research briefing: Early findings on the impact of Covid-19 on girls and young women*.

Gleason, J. (2005). *The Development of Language*. Pearson/Allyn and Bacon.

Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language*, *17*(1), 171–183.

Goldfield, B. A., Snow, C. E., & Willenberg, I. A. (2024). Variation in Language Development: Implications for Research and Theory. In J. Gleason & N. Ratner (Eds.), *The Development of Language* (10th ed.). Plural Publishing. https://www.ebsco.com/terms-of-use

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language Matters: Denying the Existence of the 30-Million-Word Gap Has Serious Consequences. In *Child Development* (Vol. 90, Issue 3, pp. 985–992). Blackwell Publishing Inc. https://doi.org/10.1111/cdev.13128

Gomez, R. (2022). *Place2Be's one-to-one counselling service in UK primary schools: an updated cost-benefit analysis*.

Goodall, J. (2013). Parental engagement to support children's learning: A six point model. *School Leadership and Management*, *33*(2), 133–150. https://doi.org/10.1080/13632434.2012.724668

Goodall, J. (2018). Leading for parental engagement: working towards partnership. In *School Leadership and Management* (Vol. 38, Issue 2, pp. 143–146). Routledge. https://doi.org/10.1080/13632434.2018.1459022

Goodall, J., Ramadan, I., Williams, M., Ylonen, A., Lewis, H., Owen, S., Roberts, D., Wolfe, C., Hughes, C., & Clegg, Z. (2021). *From Relationships to Partnerships Subtitle: The impact of the COVID-19 pandemic on parental engagement in children's learning in Wales and the implications for initial teacher education*.

Goodnow, J. J., Miller, P. J., & Kessel, F. E. (1995). *Cultural practices as contexts for development.* Jossey-Bass/Wiley.

Goodwyn, A. (2014). English and literacy in education: national policies. In *The Routledge Companion to English Studies* (pp. 16–32). Routledge.

Gore, J., Fray, L., Miller, A., Harris, J., & Taggart, W. (2021). The impact of COVID-19 on student learning in New South Wales primary schools: an empirical study. *Australian Educational Researcher*, *48*(4), 605–637. https://doi.org/10.1007/s13384-021-00436-w

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, *11*(4), 341–363.

Grabe, W., & Robert, K. (1996). *Theory and Practice of Writing*.

Graves, M. F. (1986). Chapter 2: Vocabulary learning and instruction. *Review of Research in Education*, *13*(1), 49–89.

Greenwood, C. R., Carta, J. J., Walker, D., Watson-Thompson, J., Gilkerson, J., Larson, A. L., & Schnitz, A. (2017). Conceptualizing a Public Health Prevention Intervention for Bridging the 30 Million Word Gap. In *Clinical Child and Family Psychology Review* (Vol. 20, Issue 1, pp. 3–24). Springer New York LLC. https://doi.org/10.1007/s10567-017-0223-8

Groth-Marnat, G., & Wright, A. J. (2009). *Handbook of personality assessment . Hoboken*. NJ: Wiley.

Guest, G., MacQueen, K., & Namey, E. (2014). Applied Thematic Analysis. In *Applied Thematic Analysis*. SAGE Publications, Inc. https://doi.org/10.4135/9781483384436

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, *18*(3), 218–238. https://doi.org/10.1016/j.asw.2013.05.002

Gyllstad, H., Vilkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL-International Journal of Applied Linguistics*, *166*(2), 278–306.

Halliday, M. A. K. (1978). Ideas about language. *Arts: The Journal of the Sydney University Arts Association*, *11*.

Hall-Mills, S., & Apel, K. (2015). Linguistic feature development across grades and genre in elementary writing. *Language, Speech, and Hearing Services in Schools*, *46*(3), 242–255. https://doi.org/10.1044/2015_LSHSS-14-0043

Hanushek, E. A., Light, J. D., Peterson, P. E., Talpey, L. M., & Woessmann, L. (2022). *Long-run Trends in the U.S. SES-Achievement Gap*. http://www.lbjlibrary.net/collections/selected-speeches/november-1963-1964/01-08-

Hardcastle, J., & Yandell, J. (2018). 'Even the dead will not be safe': the long war over school English. *Language and Intercultural Communication*, *18*(5), 562–575.

Harris, K. R., Graham, S., Brindle, M., & Sandmel, K. (2009). Metacognition and children's writing. *Handbook of Metacognition in Education*, 131–153.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

Hart, S. A., Little, C., & van Bergen, E. (2021). Nurture might be nature: cautionary tales and proposed solutions. In *npj Science of Learning* (Vol. 6, Issue 1). Springer Nature. https://doi.org/10.1038/s41539-020-00079-z

Heath, S. B. (1983). *Ways with words: Language, life and work in communities and classrooms*. cambridge university Press.

Hellwig, B. (2022). 2 Language acquisition and language socialization. *Approaches to Language and Culture*, *1*, 33.

Herrington, A., & Curtis, M. Smith. (2000). *Persons in process : four stories of writing and personal development in college*. National Council of Teachers of English.

Hindman, A. H., Wasik, B. A., & Snell, E. K. (2016). Closing the 30 Million Word Gap: Next Steps in Designing Research to Inform Practice. *Child Development Perspectives*, *10*(2), 134–139. https://doi.org/10.1111/cdep.12177

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). *The Difference Between Emergency Remote Teaching and Online Learning*. https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Hornby, G., & Blackwell, I. (2018). Barriers to parental involvement in education: an update. *Educational Review*, *70*(1), 109–119. https://doi.org/10.1080/00131911.2018.1388612

Howitt, D., & Cramer, D. (2007). *Introduction to research methods in psychology*. Pearson Education.

Hsiao, Y., Banerji, N., & Nation, K. (2021). Boys Write About Boys: Androcentrism in Children's Reading Experience and Its Emergence in Children's Own Writing. *Child Development*, *92*(6), 2194–2204. https://doi.org/10.1111/cdev.13623

Hsiao, Y., Dawson, N. J., Banerji, N., & Nation, K. (2022). The nature and frequency of relative clauses in the language children hear and the language children read: A developmental cross-corpus analysis of English complex grammar. *Journal of Child Language*.

Hulme, M., Beauchamp, G., Clarke, L., & Hamilton, L. (2021). Collaboration in Times of Crisis: Leading UK Schools in the Early Stages of a Pandemic. *Leadership and Policy in Schools*. https://doi.org/10.1080/15700763.2021.1917623

Hunt, P. (2006). Introduction: The expanding world of children's literature studies. In *Understanding children's literature* (pp. 1–14). Routledge.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence update? *Developmental Psychology*, *11*(6), F31–F39. https://doi.org/10.1111/j.1467-7687.2008.00768.x.Does

Huttenlocher, J., Vasilyeva, M., Waterfall, H., Vevea, J., & Hedges, L. (2007). The Varieties of Speech to Young Children. *Developmental Psychology*. https://doi.org/10.1037/0012-1649.43.5.1062.supp

Hymes, D. (1972). On Communicative Competence. *Sociolinguistics*.

Jakubowski, T. D., & Sitko-Dominik, M. M. (2021). Teachers' mental health during the first two waves of the COVID-19 pandemic in Poland. *PLoS ONE*, *16*(9 September). https://doi.org/10.1371/journal.pone.0257252

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, *19*(1), 57–84.

Jeynes, W. H. (2014). Parental Involvement That Works … Because It's Age-Appropriate. *Kappa Delta Pi Record*, *50*(2), 85–88. https://doi.org/10.1080/00228958.2014.900852

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: a develop-mental perspective. In *Phonetics Working Papers* (Vol. 53).

Johansson, Victoria. (2009). *Developmental aspects of text production in writing and speech*.

Johnson, D. C. (2012). Positioning the language policy arbiter: Governmentality and footing in the school district of Philadelphia. In *Language policies in education* (pp. 116–136). Routledge.

Johnson, D. C., & Ricento, T. (2014). Methodologies of language policy research. In *The Routledge handbook of educational linguistics* (pp. 38–49). Routledge.

Juzwik, M. M., Curcic, S., Wolbers, K., Moxley, K. D., Dimling, L. M., & Shankland, R. K. (2006). Writing into the 21st century: An overview of research on writing, 1999 to 2004. *Written Communication*, *23*(4), 451–476.

Kamimoto, T. (2008). *Nation's vocabulary levels test and its successors: A re-appraisal*.

Karami, H. (2012). The development and validation of a bilingual version of the Vocabulary Size Test. *RELC Journal*, *43*(1), 53–67.

Kartushina, N., Mani, N., Aktan-Erciyes, A., Alaslani, K., Aldrich, N. J., Almohammadi Haifa Alroqi, A., Anderson, L. M., Aussems, S., Babineau, M., Barokova, M., Bergmann, C., Cashon, C., & Custode, S. (2022). COVID-19 first lockdown as a window into language acquisition:

associations between caregiver-child activities and vocabulary gains. *Language Development Research*, *2*(1).

Kennedy, G. (2014). *An introduction to corpus linguistics*. Routledge.

Klein, P. D., & Boscolo, P. (2015). Trends in research on writing as a learning activity. *Journal of Writing Research*, *7*(3), 311–351. https://doi.org/10.17239/jowr-2016.07.03.01

Kline, P. (2000). *The handbook of psychological testing*. Psychology Press.

Knight, D., Loizides, F., Neale, S., Anthony, L., & Spasić, I. (2021). Developing computational infrastructure for the CorCenCC corpus: The National Corpus of Contemporary Welsh. *Language Resources and Evaluation*, *55*(3), 789–816. https://doi.org/10.1007/s10579-020-09501-9

Knight, D., Morris, S., Arman, L., Needs, J., & Rees, M. (2021). *Building a National Corpus: A Welsh language case study*. Springer Nature.

Knobel, M., & Lankshear, C. (2003). Researching young children's out-of-school literacy practices. *Handbook of Early Childhood Literacy*, 51–65.

Knowles, M., & Malmkjær, K. (2002). *Language and control in children's literature*. Routledge.

Kremmel, B., & Schmitt, N. (2016). Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words? *Language Assessment Quarterly*, *13*(4), 377–392.

Kuchirko, Y. (2019). On differences and deficits: A critique of the theoretical and methodological underpinnings of the word gap. *Journal of Early Childhood Literacy*, *19*(4), 533–562. https://doi.org/10.1177/1468798417747029

Kyle, K. (2019). Measuring lexical richness. In *The Routledge Handbook of Vocabulary Studies* (pp. 454–476). Taylor and Francis. https://doi.org/10.4324/9780429291586-29

Kyle, K., & Crossley, S. A. (2015). Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. *TESOL Quarterly*, *49*(4), 757–786. https://doi.org/10.1002/tesq.194

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the Validity of Lexical Diversity Indices Using Direct Judgements. *Language Assessment Quarterly*, *18*(2), 154–170. https://doi.org/10.1080/15434303.2020.1844205

Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior Research Methods*, *50*(3), 1030–1046. https://doi.org/10.3758/s13428-017-0924-4

Kyle, K., Crossley, S., & Verspoor, M. (2020). MEASURING LONGITUDINAL WRITING DEVELOPMENT USING INDICES OF SYNTACTIC COMPLEXITY AND SOPHISTICATION. *Studies in Second Language Acquisition*, 1–32. https://doi.org/10.1017/S0272263120000546

Lancaster, L. (2012). Moving into literacy: How it all begins. *The SAGE Handbook of Early Childhood Literacy*, 313–328.

Latham, D. (2002). *How children learn to write: Supporting and developing children's writing in school*. Sage.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics*, *32*(1), 69–92. https://doi.org/10.1017/S0142716410000299

Leech, G., & Rayson, P. (2014). *Word frequencies in written and spoken English: Based on the British National Corpus*. Routledge. https://books.google.co.uk/books?hl=en&lr=&id=D3DJAwAAQBAJ&oi=fnd&pg=PP1&dq=leech+rayson+wilson+2001&ots=UVpBGz2vs2&sig=nIIP5B-GuIbpowSYl3R42DVnz2Y#v=onepage&q=leech%20rayson%20wilson%202001&f=false

Letzel, V., Pozas, M., & Schneider, C. (2020). Energetic Students, Stressed Parents, and Nervous Teachers: A Comprehensive Exploration of Inclusive Homeschooling during the COVID-19 Crisis. *Open Education Studies*, *2*(1), 159–170. https://doi.org/10.1515/edu-2020-0122

Levinthal, C., Lima, O., & Kuusisto, E. (2019). Parental Engagement in Children's Learning: A Holistic Approach to Teacher-Parents' Partnerships. In K. Tirri & A. Toom (Eds.), *Pedagogy in Basic and Higher Education - Current Developments and Challenges*. www.intechopen.com

Lieberman, A., & Miller, L. (2005). Teachers as Leaders. In *The Educational Forum* • (Vol. 69).

Litosseliti, L. (2018). *Research methods in linguistics*. Bloomsbury Publishing.

Logan, J. A. R., Justice, L. M., Yumus, M., & Chaparro-Moreno, L. J. (2019). When children are not read to at home: The million word gap. *Journal of Developmental & Behavioral Pediatrics*, *40*(5), 383–386.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: evidence from a latent-variable longitudinal study. *Developmental Psychology*, *36*(5), 596.

Lorge, I., & Chall, J. (1963). Estimating the size of vocabularies of children and adults: An analysis of methodological issues. *The Journal of Experimental Education*, *32*(2), 147–157.

Lucas, M., Nelson, J., & Sims, D. (2020). *Schools' responses to Covid-19: Pupil engagement in remote learning*. https://www.nfer.ac.uk/

Luke, S. G., & Christianson, K. (2018). The Provo Corpus: A large eye-tracking corpus with predictability norms. *Behavior Research Methods*, *50*, 826–833.

Mäkinen, L., Loukusa, S., Nieminen, L., Leinonen, E., & Kunnari, S. (2014). The development of narrative productivity, syntactic complexity, referential cohesion and event content in four- to eight-year-old Finnish children. *First Language*, *34*(1), 24–42. https://doi.org/10.1177/0142723713511000

Malvern, D., & Richards, B. (2012). Measures of Lexical Richness. In *The Encyclopedia of Applied Linguistics*. Blackwell Publishing Ltd. https://doi.org/10.1002/9781405198431.wbeal0755

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development*. Springer.

Mann, S. (2011). A critical review of qualitative interviews in applied linguistics. In *Applied Linguistics* (Vol. 32, Issue 1, pp. 6–24). https://doi.org/10.1093/applin/amq043

Martin, F., & Bolliger, D. U. (2018). Engagement matters: Student perceptions on the importance of engagement strategies in the online learning environment. *Online Learning Journal*, *22*(1), 205–222. https://doi.org/10.24059/olj.v22i1.1092

Maxwell, J. A. (2012). The Importance of Qualitative Research for Causal Explanation in Education. *Qualitative Inquiry*, *18*(8), 655–661. https://doi.org/10.1177/1077800412452856

Mayor, J., & Plunkett, K. (2009). How many words do infants know, really? In *Connectionist Models Of Behaviour And Cognition II* (pp. 325–336). World Scientific.

McCallum, R., & Stephens, J. (2011). Ideology and children's books. In *Handbook of research on children's and young adult literature* (pp. 359–371). Routledge.

McCarthy, P. M., & Jarvis, S. (2007). vocd: A theoretical and empirical evaluation. *Language Testing*, *24*(4), 459–488. https://doi.org/10.1177/0265532207080767

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381

McDaniel, D., Cairns, H. S., & McKee, C. (1998). *Methods for assessing children's syntax*. Mit Press.

McEnery, T., & Wilson, A. (2003). Corpus linguistics. *The Oxford Handbook of Computational Linguistics*, 448–463.

McKee, G., Malvern, D., & Richards, B. (2000a). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, *15*(3), 323–338.

McKee, G., Malvern, D., & Richards, B. (2000b). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*, *15*(3), 323–338.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly*, 522–535.

McLean, S., Hogg, N., & Kramer, B. (2014). Estimations of Japanese university learners' English vocabulary sizes using the vocabulary size test. *Vocabulary Learning and Instruction*, *3*(2), 47–55.

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, *19*(6), 741–760.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, *27*(1), 57–86. https://doi.org/10.1177/0741088309351547

McNamara, T., & Shohamy, E. (2008). Language tests and human rights. *International Journal of Applied Linguistics*, *18*(1), 89–95.

Meara, P. (1990). Some notes on the Eurocentres vocabulary tests. *Foreign Language Comprehension and Production*, 103–113.

Meara, P., & Jones, G. (1989). *Eurocentres Vocabulary Size Test: Version E1. 1, K10, MSDOS.* Eurocentres.

Meara, P., & Milton, J. (2003). *X-lex: the Swansea levels test*. Express Publishing.

Merga, M. K., Malpique, A., Mat Roni, S., Valcan, D., & Ledger, S. (2021). Teachers' perceptions of the impact of COVID-19 on writing instruction in Australia. *Issues in Educational Research*, *31*(4), 1138.

Miller, P. J., Fung, H., Lin, S., Chen, E. C.-H., & Boldt, B. R. (2012). How socialization happens on the ground: Narrative practices as alternate socializing pathways in Taiwanese and European-American families. *Monographs of the Society for Research in Child Development*, i–140.

Mills, C., & Milton, J. (2021). Validating Pic-Lex for pre-primary and primary school age children. *Research Papers in Language Teaching and Learning*, *11*(1), 159–172. http://rpltl.eap.gr

Milton, J., & Fitzpatrick, T. (2013). *Dimensions of vocabulary knowledge*. Macmillan International Higher Education.

Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *Canadian Modern Language Review*, *63*(1), 127–147.

Milton, J., & Treffers-Daller, J. (2013). Vocabulary size revisited: The link between vocabulary size and academic achievement. *Applied Linguistics Review*, *4*(1), 151–172. https://doi.org/10.1515/applirev-2013-0007

Mind. (2020). *The mental health emergency*.

Mind. (2021). *Coronavirus: the consequences for mental health The ongoing impact of the coronavirus pandemic on people with mental health problems across England and Wales*.

Moncreieffe, M., Race, R., & Harris, R. (2020). *Decolonising the Curriculum*.

Montello, M. (2014). Narrative competence. In *Stories and their limits* (pp. 185–197). Routledge.

Morrow, L. M. (2020). *Literacy development in the early years : helping children read and write*.

Murphy, A. (n.d.). *What is Seesaw?* Retrieved January 23, 2023, from https://help.seesaw.me/hc/en-us/articles/115003713306-What-is-Seesaw-

Myhill, D. (2013). ORE Open Research Exeter TITLE From Talking to Writing: Linguistic Development in Writing A NOTE ON VERSIONS From Talking to Writing: Linguistic Development in Writing. In

British Journal of Educational Psychology Monograph Series 11 DEPOSITED IN ORE (Vol. 14). http://hdl.handle.net/10036/4478

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. *The Nature of Vocabulary Acquisition*, *19*, 35.

Nation, I. S. P., & Coxhead, A. (2021). *Measuring Native-Speaker Vocabulary Size*. John Benjamins Publishing Company.

Nation, K., Dawson, N., & Hsiao, Y. (2022). Book Language and Its Implications for Children's Language, Literacy, and Development. *Current Directions in Psychological Science*, *31*(5), 464. https://doi.org/10.1177/09637214221119448

Nation, P. (1993). Using dictionaries to estimate vocabulary size: essential, but rarely followed, procedures. *Language Testing*, *10*(1), 27–40.

Nation, P. (1995). A brief critique of Hart, B. & Risley, T.(1995). *Meaningful Differences in the Everyday Experience of Young American Children. Baltimore, MD: Paul H. Brookes. LALS, Victoria University of Wellington, New Zealand*.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*(1), 59–82.

Nation, P. (2013). *Teaching & learning vocabulary*. Boston: Heinle Cengage Learning.

Nation, P. (2016). *Making and using word lists for language learning and testing*. John Benjamins Publishing Company.

Nation, P., & Anthony, L. (2016). Measuring vocabulary size. *Handbook of Research in Second Language Teaching and Learning*, *3*, 355–368.

Nation, P., & Coxhead, A. (2014). Vocabulary size research at Victoria University of Wellington, New Zealand. *Language Teaching*, *47*(3), 398–403.

National Literacy Trust. (2020, May 5). Mitigating the significant impact of COVID-19 on the literacy of disadvantaged children | National L. *National Literacy Trust*. https://literacytrust.org.uk/blog/mitigating-significant-impact-covid-19-literacy-disadvantaged-children/

National Strategy for Educational Research and Enquiry. (2021, July 12). *Research studies on the impact of the COVID-19 pandemic on the Welsh education system*. https://hwb.gov.wales/professional-development/the-national-strategy-for-educational-research-and-enquiry-nsere/research-studies-on-the-impact-of-the-covid-19-pandemic-on-the-welsh-education-system/the-research-studies/

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining The Word Length Effect. *Psychonomic Bulletin & Review*.

New International Version. (2011). BibleGateway.com. https://www.biblegateway.com/passage/?search=Matthew+25%3A29&version=NIV

Newlove-Delgado, T., McManus, S., Sadler, K., Thandi, S., Vizard, T., Cartwright, C., & Ford, T. (2021). Child mental health in England before and during the COVID-19 lockdown. In *The Lancet Psychiatry* (Vol. 8, Issue 5, pp. 353–354). Elsevier Ltd. https://doi.org/10.1016/S2215-0366(20)30570-8

Nguyen, L. T. C., & Nation, P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal*, *42*(1), 86–99.

Niederhoff, B. (2011). Focalization. In *The Living Handbook of Narratology*. https://www.lhn.uni-hamburg.de/node/18.html

Nikolajeva, M. (2004). Narrative theory and children's literature. In *International Companian Encyclopedia of Children's Literature E2* (Second, Vol. 1). Routledge.

Nippold, M. A., Ward-Lonergan, J., & Fanning, J. L. (2005). Persuasive writing in children, adolescents, and adults: a study of syntactic, semantic, and pragmatic development Syntactic Development in Adolescents With a History of Language Impairments View project. In *Article in Language Speech and Hearing Services in Schools*. http://www.ode.state.or.us

Nocella II, A. J. (2017). Defining eco-ability. *Disability Studies and the Environmental Humanities: Toward an Eco-Crip Theory*.

Noor, S., Isa, F. Md., & Mazhar, F. F. (2020). Online Teaching Practices During the COVID-19 Pandemic. *Educational Process: International Journal*, *9*(3), 169–184. https://doi.org/10.22521/edupij.2020.93.4

O'Donnell, R. C. (1976). A critique of some indices of syntactic maturity. *Research in the Teaching of English*, *10*(1), 31–38.

Ofsted. (2020a). *COVID-19 series - briefing on schools - September 2020*.

Ofsted. (2020b). *COVID-19 series: briefing on schools, October 2020. Evidence from pilot visits to schools between 29 September and 23 October 2020*. www.gov.uk/government/publications/covid-19-series-briefing-on-schools-

Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second-and fourth-grade students. *Reading and Writing*, *22*(5), 545–565.

Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing*, *26*(1), 45–65. https://doi.org/10.1007/s11145-012-9392-5

Orlov, G., Mckee, D., Berry, J., Boyle, A., Diciccio, T., Ransom, T., Rees-Jones, A., & Stoye, J. (2020). *Learning During the COVID-19 Pandemic: It Is Not Who You Teach, but How You Teach*.

O'Sullivan, B., & Weir, C. (2011). *Language Testing: Theories and Practices*. Palgrave Macmillan.

Oxford University Press. (2018). *Why Closing The Word Gap Matters: Oxford Language Report* (O. U. Press, Ed.). Oxford University Press.

Oxford University Press. (2020). *Oxford Corpus Children's Language Report | SUMMARY 2020 | 1*.

Oxford University Press. (2023). *Building Children's Vocabulary at Home & School*.

Oxford University Press. (2024). *Oxford Corpus Children's Language Report | Analysis of BBC 500 WORDS 2023 Submissions*.

Page, L. (2021, September 24). COVID's Literacy Fallout - Language Magazine. *Language Magazine*. https://www.languagemagazine.com/2021/09/24/covids-literacy-fallout/

Parentkind. (2022). *Supporting Learning At Home*.

Paris, J., Ricardo, A., & Rymond, D. (2019). *Child Growth and Development*. College of the Canyons.

Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice*. Sage publications.

Paul, P. V. (1990). Using three test formats to assess good and poor readers' word knowledge. *Center for the Study of Reading Technical Report; No. 509*.

Pearson Assessments. (n.d.). *PPVT-4 Technical Specifications*. Retrieved March 26, 2024, from https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/ppvt-4/ppvt-4-technical-specifications.pdf

Pearson Education. (2013). *PPVT-4 Technical Report*. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/ppvt-4/ppvt-4-technical-report.pdf

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, *9*(12). https://doi.org/10.1371/journal.pone.0115844

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of Functional Literacy*, *11*, 67–86.

Pinto, G., Tarchi, C., & Bigozzi, L. (2016). Development in narrative competences from oral to written stories in five-to seven-year-old children. *Early Childhood Research Quarterly*, *36*, 1–10.

Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing*, *22*(3), 217–230. https://doi.org/10.1016/j.jslw.2013.02.003

Prior, M. T. (2018). Interviews and Focus Groups. In *The Palgrave Handbook of Applied Linguistics Research Methodology* (pp. 225–248). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-59900-1_11

Puranik, C. S., & Lonigan, C. J. (2014). Emergent Writing in Preschoolers: Preliminary Evidence for a Theoretical Framework. *Read Res Q.*, *49*(4), 453–467.

Purcell-Gates, V. (2020). *Cultural practices of literacy: Case studies of language, literacy, social practice, and power*. Routledge.

Quigley, A. (2018). *Closing The Vocabulary Gap*.

Read, J. (2000). *Assessing Vocabulary*. Cambridge University Press.

Read, J. (2004). Word Associates Test (Version 4.0). *Retrieved March*, *25*, 2015.

Reed, J., & Ort, K. (2022). The Rise of Eating Disorders During COVID-19 and the Impact on Treatment. In *Journal of the American Academy of Child and Adolescent Psychiatry* (Vol. 61, Issue 3, pp. 349–350). Elsevier Inc. https://doi.org/10.1016/j.jaac.2021.10.022

Reese, E., Suggate, S., Long, J., & Schaughency, E. (2010). Children's oral narrative and reading skills in the first 3 years of reading instruction. *Reading and Writing*, *23*, 627–644.

Rehman, R., & Fatima, S. S. (2021). An innovation in Flipped Class Room: A teaching model to facilitate synchronous and asynchronous learning during a pandemic. *Pakistan Journal of Medical Sciences*, *37*(1), 1–6. https://doi.org/10.12669/pjms.37.1.3096

Richards, B. (1987). Type/token ratios: What do they really tell us? *Journal of Child Language*, *14*(2), 201–209.

Richards, K. (2009). Interviews. In *Qualitative research in applied linguistics–A practical introduction* (pp. 182–199).

Riley, J. L. (1995). The transition phase between emergent literacy and conventional beginning reading: new research findings. *Early Years*, *16*(1), 55–59.

Roessingh, H., Elgie, S., & Kover, P. (2015). Using lexical profiling tools to investigate childrens written vocabulary in grade 3: An exploratory study. *Language Assessment Quarterly*, *12*(1), 67–86. https://doi.org/10.1080/15434303.2014.936603

Russell, D. W., Altmaier, E., & Velzen, D. Van. (1987). Job-Related Stress, Social Support, and Burnout Among Classroom Teachers. In *Journal of Applitd Plycholoty* (Issue 2).

Rychlý, P. (2008). A Lexicographer-Friendly Association Score. *RASLAN*, 6–9.

Sarland, C. (1998). The impossibility of innocence: Ideology, politics, and children's literature. In *Understanding children's literature* (pp. 49–58). Routledge.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Springer.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, *64*(4), 913–951.

Schmitt, N., Nation, P., & Kremmel, B. (2020). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching*, *53*(1), 109–120.

Schmitt, N., Ng, J. W. C., & Garras, J. (2011). The word associates format: Validation evidence. *Language Testing*, *28*(1), 105–126.

Schwenck, C. M., & Pryor, J. D. (2021). Student perspectives on camera usage to engage and connect in foundational education classes: It's time to turn your cameras on. *International Journal of Educational Research Open*, *2*. https://doi.org/10.1016/j.ijedro.2021.100079

Seesaw. (n.d.). *Seesaw For Schools efficacy study*. https://goo.gl/f2TwlX

Segbers, J., & Schroeder, S. (2017). How many words do children know? A corpus-based estimation of children's total vocabulary size. *Language Testing*, *34*(3), 297–320. https://doi.org/10.1177/0265532216641152

Seginer, R. (2006). *Parents' Educational Involvement: A Developmental Ecology Perspective*.

Seidenberg, M. S., & MacDonald, M. C. (2018). The impact of language experience on language and reading a statistical learning approach. *Topics in Language Disorders*, *38*(1), 66–83. https://doi.org/10.1097/TLD.0000000000000144

Sénéchal, M., Ouellette, G., & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. *Handbook of Early Literacy Research*, *2*, 173–182.

Senedd. (2021). *The impact of COVID-19 on children and young people: Final report*. www.senedd.wales

Serafino, P. (2019). *Exploring the UK's digital divide*.

Sharp, C., Nelson, J., Lucas, M., Julius, J., Mccrone, T., & Sims, D. (2020). *Schools' responses to The challenges facing schools and pupils in September 2020*. https://www.nfer.ac.uk/

Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, *18*(4), 373–391.

Shohamy, E. (2005). Language policy: Hidden agendas and new approaches. In *Language Policy: Hidden Agendas and New Approaches*. Routledge Taylor & Francis Group. https://doi.org/10.4324/9780203387962

Sibieta, L. (2021, February 1). *The crisis in lost learning calls for a massive national policy response - Institute For Fiscal Stud*. Institute for Fiscal Studies . https://ifs.org.uk/publications/15291

Sibieta, L., & Cottell, J. (2020). *Education policy responses across the UK to the pandemic*. www.nuffieldfoundation.org

Silverman, D. (2015). *Interpreting qualitative data*. Sage.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press, USA.

Sinclair, J. (2005). Meaning in the framework of corpus linguistics. *Lexicographica (2004)*, *20*, 20–32.

Smith, N., & McEnery, T. (1998). Issues in Transcribing a Corpus of Children's Hanwritten Projects. *Literary and Linguistic Computing*, *13*(4), 217–225.

Sperry, D. E., Sperry, L. L., & Miller, P. J. (2019a). Language Does Matter: But There is More to Language Than Vocabulary and Directed Speech. In *Child Development* (Vol. 90, Issue 3, pp. 993–997). Blackwell Publishing Inc. https://doi.org/10.1111/cdev.13125

Sperry, D. E., Sperry, L. L., & Miller, P. J. (2019b). Reexamining the Verbal Environments of Children From Different Socioeconomic Backgrounds. *Child Development*, *90*(4), 1303–1318. https://doi.org/10.1111/cdev.13072

Spielman, A. (2020, November 22). *HMCI commentary: findings from visits in October - GOV*. Ofsted. https://www.gov.uk/government/speeches/hmci-commentary-findings-from-visits-in-october

Stadler, M., & Ward, G. (2006). Supporting the narrative development of young children. *Early Childhood Education Journal*, *33*(2), 73–80. https://doi.org/10.1007/s10643-005-0024-4

Steinberg, D., & Sciarini, N. (2006). *An Introduction to Psycholinguistics* (2nd ed.). Pearson Education.

Stephens, J. (2005). Linguistics and Stylistics. In *Understanding Children's Literature*. http://ebookcentral.proquest.com

Stewart, J. (2014). Do multiple-choice options inflate estimates of vocabulary size on the VST? *Language Assessment Quarterly*, *11*(3), 271–282.

Stewart, J., & White, D. A. (2011). Estimating guessing effects on the vocabulary levels test for differing degrees of word knowledge. *Tesol Quarterly*, *45*(2), 370–380.

Stockman, I. J. (2000). The new Peabody Picture Vocabulary Test—III: an illusion of unbiased assessment? *Language, Speech, and Hearing Services in Schools*, *31*(4), 340–353.

Street, B. (1984). *Literacy in theory and practice* (Vol. 9). Cambridge University Press.

Sullivan, J. R., Winter, S. M., Sass, D. A., & Svenkerud, N. (2014). Assessing growth in young children: A comparison of raw, age-equivalent, and standard scores using the peabody picture vocabulary test. *Journal of Research in Childhood Education*, *28*(2), 277–291. https://doi.org/10.1080/02568543.2014.883453

Sulzby, E., & Teale, W. H. (1985). Writing development in early childhood. *Educational Horizons*, *64*(1), 8–12.

Sun, A., & Chen, X. (2016). Online Education and Its Effective Practice: A Research Review. In *Journal of Information Technology Education: Research* (Vol. 15). http://www.informingscience.org/Publications/3502

Sun, L., & Nippold, M. A. (2012). Narrative writing in children and adolescents: Examining the literate lexicon. *Language, Speech, and Hearing Services in Schools*, *43*(1), 2–13. https://doi.org/10.1044/0161-1461(2011/10-0099)

Sun, Y., Zhang, J., & Scardamalia, M. (2010). Knowledge building and vocabulary growth over two years, Grades 3 and 4. *Educational Theory and Practice*, *10*. http://scholarsarchive.library.albany.edu/etap_fac_scholarhttp://scholarsarchive.library.albany.edu/etap_fac_scholar/10

Talmy, S. (2010). Qualitative interviews in applied linguistics: From research instrument to social practice. In *Annual Review of Applied Linguistics* (Vol. 30, pp. 128–148). https://doi.org/10.1017/S0267190510000085

Thompson, P., & Sealey, A. (2007). Through children's eyes?: Corpus evidence of the features of children's literature. *International Journal of Corpus Linguistics*, *12*(1), 1–23. https://doi.org/10.1075/ijcl.12.1.03tho

Timmins, N. (2021). *Schools and coronavirus: The government's handling of the education during the pandemic*. www.instituteforgovernment.org.uk

Turner, C. (2015, October 27). *No Child Left Behind: What Worked, What Didn't*. NPR. https://www.npr.org/sections/ed/2015/10/27/443110755/no-child-left-behind-what-worked-what-didnt

Turner, D. W. (2010). Qualitative Interview Design: A Practical Guide for Novice Investigators. In *The Qualitative Report* (Vol. 15). http://www.nova.edu/ssss/QR/QR15-3/qid.pdf

Uccelli, P., Dobbs, C. L., & Scott, J. (2013). Mastering Academic Language: Organization and Stance in the Persuasive Writing of High School Students. *Written Communication*, *30*(1), 36–62. https://doi.org/10.1177/0741088312469013

Underwood, P., Brown, N., Te Riele, K., Shelley, B., Woodroffe, J., Bakhtiar, A., Crellin, T., Ellis, S., Grant, O., Iwanicki, V., Johnston, G., Koirala, S., Lang, M., Mundey, D., Stewart, S., & West, M. (2020). *Learning at home during COVID-19: Effects on vulnerable young Australians Independent Rapid Response Report*. http://creativecommons.org/licenses/by/3.0/au/deed.en.Thefulllicencetermsareavailablefrom http://creativecommons.org/licenses/by/3.0/au/legalcode.

Van Hout, R., & Vermeer, A. (2007). Comparing measures of lexical richness. *Modelling and Assessing Vocabulary Knowledge*, *93*, 115.

Van Lancker, W., & Parolin, Z. (2020). COVID-19, school closures, and child poverty: a social crisis in the making. In *The Lancet Public Health* (Vol. 5, Issue 5, pp. e243–e244). Elsevier Ltd. https://doi.org/10.1016/S2468-2667(20)30084-0

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, *22*(2), 217–234.

Wall, B., & Crevecoeur, Y. (2016). *Narrator's Voice: The Dilemma Of Children's Fiction*. Springer.

Waters-Davies, J., Davies, P., Underwood, C., Lloyd-Williams, S., Jones, S., James, S., Thomas-Walters, J., Morgan, C., Southern, A., Ward, A., & Aled Evans, R. (2021). *Exploring the impact of the Covid-19 pandemic on learners in Wales*.

Webb, S. (2021). WORD FAMILIES AND LEMMAS, NOT A REAL DILEMMA. In *Studies in Second Language Acquisition* (Vol. 43, Issue 5, pp. 973–984). Cambridge University Press. https://doi.org/10.1017/S0272263121000760

Weir, C. J. (2005). Language testing and validation. *Hampshire: Palgrave McMillan*.

Wellman, R. L., Lewis, B. A., Freebairn, L. A., Avrich, A. A., Hansen, A. J., & Stein, C. M. (2011). *Narrative ability of children with speech sound disorders and the prediction of later literacy skills*.

Wells, G. (2009). The social context of language and literacy development. *Handbook of Child Development and Early Education: Research to Practice*, 271–302.

Welsh Government. (2020). *Assessing the impact of the initial policy decisions taken in relation to the provision of education in response to Covid-19*. https://gov.wales/statement-minster-education-kirsty-williams-school-closures-wales

Welsh Government. (2021, May 20). *Timeline of school closures during the coronavirus (COVID-19) pandemic, March 2020 to April 2021*. Welsh Government . https://gov.wales/timeline-school-closures-during-coronavirus-covid-19-pandemic

Wheeler, D. L., & Hill, J. C. (2021). The impact of COVID-19 on early childhood reading practices. *Journal of Early Childhood Literacy*. https://doi.org/10.1177/14687984211044187

White, A. (2013, April 25). *Four reasons why policy-making shouldn't be outsourced to right-wing think tanks*. The New Statesman. https://www.newstatesman.com/politics/2013/04/four-reasons-why-policy-making-shouldnt-be-outsourced-right-wing-think-tanks

White, T. G., Graves, M. F., & Slater, W. H. (1990). Growth of Reading Vocabulary in Diverse Elementary Schools: Decoding and Word Meaning. *Journal of Educational Psychology*, *82*(2), 281–290. https://doi.org/10.1037/0022-0663.82.2.281

Whitehead, M. R. (2010). *Language & literacy in the early years 0-7*. Sage.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, *69*(3), 848–872.

Widnall, E., Winstone, L., Mars, B., Haworth, C., & Kidger, J. (2020). *Initial findings from a secondary school survey study in South West England. Young People's Mental Health during the COVID-19 Pandemic*.

Wild, K., Kilgarriff, A., & Tugwell, D. (2013). The Oxford Children's Corpus: Using a Children's Corpus in Lexicography. *International Journal of Lexicography*, *26*(2), 190–218. https://doi.org/10.1093/ijl/ecs017

Wodak, R. (2008). Introduction: Discourse Studies – Important Concepts and Terms. In R. Wodak & M. Krzyżanowski (Eds.), *Qualitative Discourse Analysis in the Social Sciences* (pp. 1–24). Palgrave Macmillan.

Woodcock, A. (2021, March 23). 90,000 children 'lacking basic literacy' because of Covid, Boris Johnson warns | The Independent. *The Independent*. https://www.independent.co.uk/news/uk/politics/boris-johnson-coronavirus-children-literacy-b1821353.html

Yandell, J. (2017). Culture, Knowledge and Power: What the Conservatives have Learnt from E.D. Hirsch. *Changing English: Studies in Culture and Education*, *24*(3), 246–252. https://doi.org/10.1080/1358684X.2017.1351231

Zegiob, L. E., Arnold, S., & Forehand, R. (1975). *An Examination of Observer Effects in Parent-Child Interactions* (Vol. 46, Issue 2).

Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing*, *47*. https://doi.org/10.1016/j.asw.2020.100505

9. Appendices

9.1 Appendix 1. Picture Vocabulary Size Test Set #1

Appendix 1. PVST Test Set 1

101011 Behind: He's behind the car.

101021 By: He's by the car.

101031 Thirteen: Thirteen.

101041 House: This is my house.

101051 Wild: It's wild.

101061 Animal: It's an animal.

101071 Table: It's a table.

101081 Grass: It's grass.

102091 Message: It's a message.

102101 Attack: It's an attack.

102111 Lake: It's a lake.

102121 Afraid: It's afraid.

102131 Breath: He takes a breath.

102141 Believe: He believes me.

102151 Cream: Some cream.

102161 Beneath: You can see from beneath.

103171 Whip: It's a whip.

103181 Handkerchief: It's a handkerchief.

103191 Check: It was checked.

103201 Beast: It's a beast.

103211 Knowledge: He has a lot of knowledge.

103221 Earn: He earned it.

103231 Mail: He has some mail.

103241 Frame: He has a frame.

104251 Tour: He's on a tour.

104261 Video: He has a video.

104271 Various: He has various things.

104281 Penalty: He has a penalty.

104291 Hobby: It is a hobby.

104301 Hotel: This is a good hotel.

104311 Tape: This is a tape.

104321 Electric: This is electric.

105331 Beef: This is beef.

105341 Gear: This is my gear.

105351 Liquid: This is a liquid.

105361 Award: This is an award.

105371 Independent: He is doing it independently.

105381 Investigate: It's investigating.

105391 Display: He can see the display.

105401 Adopt: He is adopted.

106411 Flock: It's a flock.

106421 Calf: It's a calf.

106431 Laundry: It's the laundry.

106441 Function: It's not functioning.

106451 Confirm: She confirmed it.

106461 Object: She objected.

106471 Alert: She's very alert.

106481 Horizon: She saw the horizon.

107491 Bully: He is being a bully.

107501 Signature: He is making his signature.

107511 Rotate: He is rotating it.

107521 Thrust: He's making a thrust.

107531 Cafeteria: This is a cafeteria.

107541 Cushion: This is a cushion.

107551 Chap: This is a good chap.

107561 Limb: This is a limb.

108571 Grasshopper: It's a grasshopper.

108581 Quaint: It's quaint.

108591 Compass: It's a compass.

108601 Savage: It's savage.

108611 Expedition: It's an expedition.

108621 Shabby: He is shabby.

108631 Chant: He's chanting.

108641 Kite: He's got a kite.

109651 Portable: It's portable.

109661 Lunar: It's lunar.

109671 Slick: There's a slick.

109681 Fatal: It's fatal.

109691 Sloppy: It' s sloppy.

109701 Merit: It has merit.

109711 Scandal: It was a scandal.

109721 Thistle: It's a thistle.

110731 Spa: This is a spa.

110741 Pulley: This is a pulley.

110751 Canary: This is a canary.

110761 Jig: This is a jig

110771 Reap: He is reaping.

110781 Rinse: He is rinsing.

110791 Trample: He is being trampled.

110801 Enhance: He is enhancing it.

111811 Crimson: It's crimson.

111821 Smudge: It's smudged.

111831 Sleet: It's sleet.

111841 Gospel: It's the gospel.

111851 Souvenir: It is a souvenir.

111861 Obnoxious: It's obnoxious.

111871 Anguish: Full of anguish.

111881 Sag: It is sagging.

112891 Brachiosaur: A brachiosaur.

112901 Disillusioned: Disillusioned.

112911 Goalie: A goalie.

112921 Licorice: Licorice.

112931 Pansy: It's a pansy.

112941 Sardine: It's a sardine.

112951 Volley: It's a volley.

112961 Stupendous: It's stupendous.

Return to location

9.2 Appendix 2. Picture Vocabulary Size Test Set #2

201011 Along: Along there.

201021 Short: It's short.

201031 Near: It's near.

201041 Many: Many.

201051 Come: She's coming.

201061 Read: She's reading.

201071 Reach: She's reaching.

201081 Quiet: She's quiet.

202091 Roof: It's a roof.

202101 Telephone: It's a telephone.

202111 Shell: It's a shell.

202121 Mud: It's mud.

202131 Sort: She's sorting.

202141 Agree: She agreed.

202151 West: She went west.

202161 Tiny: It's tiny.

203171 Occasion: This is an occasion.

203181 Couple: This is a couple.

203191 Knight: This is a knight.

203201 Nest: This is a nest.

203211 Treat: It's a treat.

203221 Direct: She's directing them.

203231 Perch: It's perching.

203241 Waste: It's waste.

204251 Threat: There's a threat

204261 Detail: There's detail.

204271 Individual: This is an individual.

204281 Reunion: There's a reunion.

204291 Nope: Nope.

204301 Shame: He felt shame.

204311 Grade: He got a grade.

204321 Credit: He has credit.

205331 Weigh: He's weighing it.

205341 Crisp: It's crisp.

205351 Promote: He was promoted.

205361 Awkward: He's awkward.

205371 Carol: A carol.

205381 Method: A method.

205391 Gang: A gang.

205401 Fort: A fort.

206411 Sticky: It's sticky.

206421 Assault: It's an assault.

206431 Choir: She's in a choir.

206441 Vow: She's making a vow.

206451 Reflection: It's a reflection.

206461 Chore: This is a chore.

206471 Intelligent: It's highly intelligent.

206481 Loaf: This is a loaf.

207491 Caravan: It's a caravan.

207501 Feast: It's a feast.

207511 Sponge: It's a sponge.

207521 Tissue: It's a tissue.

207531 Mend: He is mending it.

207541 Greet: He's greeting them.

207551 Subscribe: He's subscribing to it.

207561 Triangle: He's making a triangle.

208571 Lark: It's a lark.

208581 Moose: It's a moose.

208591 Moron: It's a moron.

208601 Bandage: It's a bandage.

208611 Batter: It's battered.

208621 Magnificent: It's magnificent.

208631 Topple: It's toppling.

208641 Crouch: It's crouching.

209651 Conquest: It's a conquest.

209661 Blouse: It's a blouse.

209671 Jug: It's a jug.

209681 Impulse: It's an impulse.

209691 Longitude: It shows longitude.

209701 Enthusiasm: It shows enthusiasm.

209711 Flit: It's flitting.

209721 Coordinate: It's being coordinated.

210731 Drought: There's a drought.

210741 Agenda: There's an agenda.

210751 Outcome: This is the outcome.

210761 Mascot: This is the mascot.

210771 Exterior: It's the exterior.

210781 Opponent: It's an opponent.

210791 Substantial: It's substantial.

210801 Hyacinth: It's a hyacinth.

211811 Cicada: It's a cicada.

211821 Mystical: It's mystical.

211831 Margin: It's the margin.

211841 Fluke: It's a fluke.

211851 Recite: She's reciting.

211861 Tactful: She's tactful.

211871 Ambitious: She's ambitious.

211881 Lull: She's lulling it.

212891 Banjo: It's a banjo.

212901 Mite: It's a mite.

212911 Pterodactyl: It's a pterodactyl.

212921 Snowdrop: It's a snowdrop.

212931 Clank: There's clanking.

212941 Ferocious: It's ferocious.

212951 Icing: It's icing.

212961 Togs: Someone's togs.

Return to location

9.3 Appendix 3. Original Project Summary

Vocabulary is one of the key predictors of academic success and vocabulary deficits can limit a child's educational development. The 'word gap' has been a pervasive issue in vocabulary research since the nineties and remains an issue today. However, our understanding of these issues is limited by a lack of normative figures for vocabulary growth and development in the early years. Currently few vocabulary tests exist that can give an accurate vocabulary size for L1 English children, especially for those who are pre-literate and so the vocabularies of younger school-age children are underreported.

This research seeks to answer the following principal questions:

i) How much vocabulary and which vocabulary is learned each year from three years' old up to the age of eleven?

ii) Is there evidence from OUP's Learner Corpus, which contains data over 20 years, that vocabulary knowledge is diminishing?

iii) How is the lexical acquisition observed in this project best explained through theories and models of lexical acquisition in young learners?

These questions will be addressed through several experiments. The vocabulary size of a number of school-age children will be measured using a receptive vocabulary measure based on and/or adapted from prior vocabulary tests such as X-Lex, Pic-Lex, the BPVS, etc. The test will have to be sensitive enough to measure the variation in scores which underlies variation in educational performance, and also identify lexical items that distinguish high performing learners from low performing learners. Repeated experiments on vocabulary will be carried out in an effort to investigate longitudinally how the lexicon develops in school-age children.

The second stage of this research project will be to use Oxford University Press' corpus of child English to examine whether lexical knowledge has diminished over the last 20 years. Changes in lexical sophistication will be examined and if the number and range of words used by children is diminishing over time then this will support the hypothesis that overall vocabulary knowledge is also diminishing. This data will be contextualised in the literature regarding the theories and models of lexical acquisition in children.