DEPOSIT AND CONSULTATION OF DISSERTATION

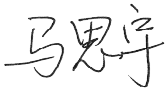
One copy of your dissertation will be deposited in ORA (Oxford University Research Archive), where it is intended to be freely available online. In order to facilitate this, you are requested to complete and sign the form below.

Please use block capitals

Surname MA
First names (in full) SIYU
Faculty board EDUCATION
Degree name and pathway MSC IN EDUCATION (RESEARCH DESIGN AND METHODOLOGY)
Title of dissertation FAMILY BACKGROUND AND HOME LITERACY ENVIRONMENT AS PREDICTORS OF THE EARLY LITERACY DEVELOPMENT OF CHILDREN IN WALES – FINDINGS FROM THE MILLENNIUM COHORT STUDY N.B. The title stated here must be precisely the same as that stated on the title page of the thesis submitted. **A candidate wishing to amend the title previously approved by the faculty must apply to the faculty board for permission to do so.**
Supervisor DR. SONALI NAG
Subject keywords *Enter your own keywords or phrases to describe your work. This information helps us describe your work in ORA* A SECONDARY DATA ANALYSIS ON THE MILLENNIUM COHORT STUDY DATASET, INVESTIGATING THE DEVELOPMENTAL RELATIONSHIPS BETWEEN FAMILY BACKGROUND, HOME LITERACY ENVIRONMENT, AND WELSH CHILDREN'S EARLY ENGLISH LITERACY AND LANGUAGE DEVELOPMENT, USING A LONGITUDINAL STRUCTURAL EQUATION MODEL
Research methods used *This information helps us describe your work on SOLO for future students e.g. quantitative, interviews, vocabulary test, systematic review, etc.* SECONDARY DATA ANALYSIS, ADVANCED QUANTITATIVE METHODS, LONGITUDINAL STRUCTURAL EQUATION MODELLING

Declaration by the candidate as author of the dissertation

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.
2. I understand that the Department requires that I shall deposit one copy of my dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].
3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals, and that such material should be removed before this dissertation is deposited in ORA.
4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [www.bodleian.ox.ac.uk/ora/deposit-in-ora/deposit-licence] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.
5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]: 

Date: 27 August 2021

Family Background and Home Literacy Environment as Predictors of the Early Literacy Development of Children in Wales - Findings from the Millennium Cohort Study

Siyu Ma

Department of Education, University of Oxford

Master of Science in Education (Research Design and Methodology)

Dr Sonali Nag

27 August, 2021

Abstract

Early literacy and language development plays a vital role in later school success. Home literacy environment has long been recognised as a strong predictor of such development, which is in turn influenced by a child's family background. We aimed to replicate some of these well-established relationships in a sample of 612 children in Wales, with secondary data obtained from a UK national longitudinal survey, the Millennium Cohort Study. Specifically, we examined the developmental relationships between family background around childbirth (family income, maternal education, maternal literacy/numeracy skills), home literacy environment at age 3 (home tutoring, home literacy practices), precursor English literacy and language skills at age 5 (vocabulary, phonological awareness, print concept), and later English literacy and language outcomes at age 7 (decoding, speaking/listening, reading, writing), controlling for child sex, season of birth, and household single-parent status. Results from a longitudinal structural equation model showed that the effects of family background on literacy and language development were partially mediated by variations in home literacy practices, but not home tutoring; meanwhile, effects of home literacy practices (but not home tutoring) on later outcomes at age 7 were fully mediated by the precursor skills at age 5. Additionally, direct pathways were revealed between family income to early vocabulary at age 5, between maternal education to English writing at age 7, and between maternal literacy/numeracy skills and English reading at age 7. The results suggest that home tutoring may be insufficient to close the attainment gaps from early disparities, and the effects of family background are more salient than previously thought.

Acknowledgements

First and foremost, I would like to express my deep and sincere thanks to my dissertation supervisor, Dr Sonali Nag. Without her support, guidance, and dedicated involvement throughout my project, this paper would never have been accomplished. I would like to thank her for all the early-morning and late-night meetings, comments on my dissertation drafts, and for offering me the opportunity to engage in the TalkTogether project as an intern during Trinity Term.

Apart from my supervisor, I would like to thank Prof Lars-Erik Malmberg and Dr Ariel Lindorff for providing generous and valuable guidance and suggestions on my data analysis. Learning to run a longitudinal structural equation model has been a difficult task, without the Advanced Quantitative Methods Summer School led by Prof Lars-Erik Malmberg, my statistical analysis would not have been successful.

I would like to extend my sincere gratitude to the Department of Education at the University of Oxford, and to all the staff in the Department who have helped me in any way. I would also like to thank my cohort peers, who have been inspirational and encouraging. I have been learning from them, from the conversations and discussions we had, and I believe I will treasure these memories for a lifetime.

I am extremely grateful to my family, especially my parents and my 4-year-old sister for their love and support during my education. I am very much thankful to all my friends in Student Castle Oxford, as well as my friends' friends, whom I played board games and celebrated birthdays with. This work would not have been possible without their support and encouragement.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background

Children's literacy and language skills in the early school years build a foundation for later literacy and language development as well as academic achievement (Duncan et al., 2007; Hamilton et al., 2016). At school entry, children bring a diverse range of skills at different levels to formal reading instruction, including oral language, phonological awareness, and print concept (Sénéchal et al., 2001). Prior to school entry, however, these early skills are largely shaped by the home literacy environment (HLE) children experience (Foster et al., 2005; Frijters et al., 2000; Sénéchal & LeFevre, 2002), which in turn is associated with a child's family background, such as mother's educational level (Bornstein et al., 2003; Ermisch & Francesconi, 2001). Indeed, the developmental relationships among family background, HLE, and children's literacy and language development have been researched extensively (e.g., Hamilton et al., 2016; Hood et al., 2008; Inoue et al., 2008).

This study aims to replicate some of these relationships, especially by referring to the theoretical model developed in Hamilton et al.'s (2016) study in the UK (details in Section 2.5). We focus on a child sample in Wales, using data from a UK national longitudinal survey, the Millennium Cohort Study (MCS). Specifically, we examine whether family background around childbirth (family income, maternal education, mother's self-reported literacy/numeracy skills) and HLE at age 3 (home tutoring, home literacy practices) predict Welsh children's English literacy and language skills measured at ages 5 (vocabulary, phonological awareness, print concept) and 7 (decoding, speaking/listening, reading, writing). We also link the precursor skills at age 5 to later literacy and language outcomes at age 7.

The child population in Wales is relatively under-researched in terms of the predictors of English literacy development. Wales is identified as a stable bilingual community, as children are exposed to English and Welsh from an early age, although English is relatively dominant (Gathercole & Thomas, 2009). This unique language profile of Wales motivates this study, to reveal whether Welsh children's English literacy and language development is comparable to that of English monolinguals. Our findings may inform the design of family training programmes to promote Welsh children's English competencies, which may bring longer-term benefits if these children wish to pursue higher education or employment beyond the Welsh-speaking contexts (Gathercole & Thomas, 2009).

Apart from this unique sample, we focus on a broader range of HLE measures than most previous studies. The home literacy model proposed by Sénéchal and LeFevre (2002) has been influential in this field, which classifies HLE into formal and informal interactions. However, the interactions specified by this model are limited to those related to reading and writing only, namely, formal HLE was operationalised as literacy instruction, and informal HLE as storybook exposure (e.g., Hamilton et al., 2016; Sénéchal & LeFevre, 2002). We argue that HLE should be conceptualised to encompass more attributes of the home that can potentially facilitate literacy and language development (Nag et al., 2019), such as watching television and teaching number counting (Purpura & Napoli, 2015; Van Steensel, 2006). Therefore, we use different terms to represent the HLE measured in our study, with home tutoring representing formal HLE, and home literacy practices representing informal HLE (see Section 2.3.2).

1.2 The Construct of Emergent Literacy

In early literacy development research, a key question concerns what constitutes the construct of emergent literacy. Indeed, there have been some theoretical frameworks classifying emergent-literacy behaviours. For example, Mason and Stewart (1990) proposed four components of emergent literacy, consisting of concepts and functions of literacy (e.g., children's knowledge on the use of newspapers), writing and composing (e.g., word, sentence writing), knowledge about letters and words (e.g., word recognition skills, phonological awareness), and comprehension and word understanding (e.g., children's ability to define words and understand narrative). Whitehurst and Lonigan (1998), however, classified emergent literacy into two broader components that were presumed to be independent of each other – inside-out and outside-in skills. The inside-out component represents children's understanding of how the print translates into sounds, including letter-name and letter-sound knowledge, punctuation, and sentence grammar knowledge, as well as meta-linguistic skills like phonological awareness. Conversely, the outside-in component refers to children's understanding of the narrative, conceptual, and semantic context of a piece of writing, including skills like vocabulary, narrative construction, print concept, and reading comprehension (Whitehurst & Lonigan, 1998).

Sénéchal et al. (2001) argued that emergent-literacy behaviours identified in the two theoretical frameworks above can be re-classified into literacy skills (i.e., conceptual knowledge about literacy, procedural knowledge about reading and writing), oral language

skills (e.g., vocabulary, narrative construction) and meta-linguistic skills (i.e., awareness of language structure, such as phonological awareness). Regarding Mason and Stewart's (1990) framework, the first and second components can tap the conceptual and procedural knowledge about literacy respectively; the third component may be aligned with meta-linguistic skills; and the fourth component appears identifiable to oral language skills. Similarly, in Whitehurst and Lonigan's (1998) framework, the inside-out component includes conceptual knowledge about literacy (e.g., letter knowledge, grammar), and meta-linguistic skills, whereas the outside-in component encompasses both conceptual (e.g., print concept) and procedural (e.g., reading comprehension) knowledge about literacy and oral language skills (e.g., vocabulary). In this study, we follow Sénéchal et al.'s (2001) framework of oral language, literacy, and meta-linguistic skills.

1.3 The Current Study

This study will replicate and extend Hamilton et al.'s (2016) model, and investigate the developmental relationships between family background, HLE, and English literacy and language development among a sample of children in Wales. This will be achieved via a secondary data analysis on the MCS, a longitudinal survey following the development of originally over 19,000 children (Hansen, 2014). Its longitudinal nature allows the developmental relationships to be investigated (MacCallum & Austin, 2000). Meanwhile, the MCS contains data on socioeconomic status, parent reports on childcare, cohort members' performance on cognitive assessments and teacher reports on children's abilities, covering the range of variables needed for this study.

Additionally, some latent constructs will be created to capture several variables of interest. Latent constructs are commonly used to represent abstract constructs that are thought to exist but cannot be observed directly (Bollen & Lennox, 1991). An example of this is general intelligence (e.g., McGrew, 2009; Panizzon et al., 2014), where people's intelligence can only be observed via their performance on various cognitive tasks. Also, the performance on any task is conceptualised to be affected by an underlying common factor, the general intelligence (Bollen & Lennox, 1991). In this study, mother's self-reported literacy/numeracy skills around childbirth, and the two components of HLE, home tutoring and home literacy practices at age 3, will be treated as latent variables, while other variables will be analysed as observed variables. Consequently, the statistical technique of structural equation modelling will be adopted to answer the three research questions listed below, as it can estimate the

relationships among observed and latent variables (Hoyle, 1995; MacCallum & Austin, 2000).

1.4 Research Questions

1. How does family background around childbirth relate to HLE at age 3, to early English literacy and language skills at age 5, and to later English literacy and language outcomes at age 7 for children in Wales?

2. How does HLE at age 3 predict the precursor skills at age 5 and later literacy and language outcomes at age 7?

3. How do early literacy and language skills at age 5 predict later literacy and language outcomes at age 7?

1.5 Structure of Paper

This paper consists of five chapters. The introduction chapter describes the aims of this study, providing background information on key constructs, sample, analytic approach, and research questions. The literature review chapter examines the existing literature related to family background, HLE, and children's early literacy and language development, by introducing each construct and their relationships. The chapter also explains what the current study wants to replicate and add to the literature. The theoretical model adopted from Hamilton et al.'s (2016) study, research hypotheses, and the hypothesised model are presented. The methodological chapter outlines characteristics of the MCS database, the analytic sample, and ethical considerations. The measurements involved in the analysis are presented, followed by an overview of the data analysis procedure, including the statistical techniques considered. The results chapter presents the findings of the statistical analyses, specifically, the measurement model, structural model, and mediation analyses. Finally, the discussion chapter interprets the statistical findings and analyses how they fit in with the research questions and preceding literature. The dissertation concludes with a discussion on strengths, limitations, and possible implications.

Chapter 2: Literature Review

This chapter will review previous literature about family background, home literacy environment (HLE) and children's early literacy and language development. Firstly, three constructs of literacy will be discussed, before reviewing the concurrent and longitudinal relationships among the literacy skills. Secondly, a broader range of distal family background and proximal HLE variables compared to what was commonly used in previous research will be proposed. Thirdly, studies on the relationships between family background, HLE, and children's early literacy and language development will be examined. The theoretical model adopted from Hamilton et al. (2016) will be presented, followed by an overview of the research hypotheses and the hypothesised model.

2.1 Early Literacy and Language Development

The acquisition of literacy has been conceptualised as a developmental continuum starting from the early years of life (Teale & Sulzby, 1986; Whitehurst & Lonigan, 1998). A variety of preschool literacy skills developed by children can lay the foundation for later reading and writing, and thus school achievement (e.g., Hood et al., 2008; Sénéchal et al., 2001; Storch & Whitehurst, 2002).

2.1.1 Literacy, Language, and Meta-linguistic Skills as Three Separate Constructs

According to Mason and Stewart (1990) and Whitehurst and Lonigan (1998), emergent literacy is a unitary construct with several components. However, Sénéchal et al. (2001) argued that emergent literacy should be treated as three constructs (i.e., literacy, oral language, meta-linguistic skills) that are highly interrelated but distinct from each other.

There has been strong evidence differentiating literacy and language as they seem to follow different developmental pathways (Goswami, 2001; Jordan et al., 2000). From an evolutionary perspective, the development of language precedes that of literacy (Geary, 1995). Specifically, language develops naturally as infants interact with the environment, while the acquisition of literacy is 'biologically secondary' (Geary, 1995, p. 24) and usually requires specific instruction. Furthermore, empirical research has demonstrated that all children acquire oral language skills to some extent prior to formal schooling, whereas literacy skills tend not to be acquired in preschool (Jordan et al., 2000). Thus, there is

seemingly a sequential order in language and literacy development, meaning that it may be more sensible to treat them as separate constructs.

Additionally, previous literature supports the distinction between literacy, language, and meta-linguistic skills, by statistically assessing whether they formed a single construct. For instance, Whitehurst et al.'s (1994) principal component analysis revealed that measures of oral language, writing (a procedural literacy skill), and meta-linguistic skills loaded on different factors. Similarly, Lonigan et al. (2000) conducted confirmatory factor analyses, showing that models with print concept (a conceptual literacy skill), oral language, and phonological awareness (a meta-linguistic skill) as separate constructs yielded better model fit than models with a single emergent-literacy construct. Therefore, statistical analyses support the notion that emergent literacy is not a unitary construct.

In practice, many researchers have analysed the three constructs separately. To exemplify, Lee and Al Otaiba's (2015) made a priori decisions to analyse literacy (e.g., alphabet knowledge), language (e.g., word spelling), and meta-linguistic (e.g., phonological awareness) skills separately in their study on socio-economic and gender differences in early literacy. Moreover, some researchers have investigated the effectiveness of literacy interventions based on a priori decisions to separate print concept, vocabulary, and phonological awareness (e.g., Chaney, 1994; Lefebvre et al., 2011; Shamir et al., 2012). Hence, the separation of literacy, oral language, and meta-linguistic skills has been repeatedly adopted in empirical practices.

Accordingly, this dissertation followed previous practices and treated literacy, oral language, and meta-linguistic skills as separate constructs when examining early literacy. This is because the three constructs have been suggested to have different developmental pathways (e.g., Jordan et al., 2000), to load on distinct factors in statistical analyses (e.g., Whitehurst et al., 1994), and to be assessed separately in previous empirical studies (e.g., Lee & Al Otaiba, 2015). Particularly, we examined early literacy in terms of vocabulary, phonological awareness, and print concept. As a meta-linguistic skill, phonological awareness is defined as the implicit and explicit sensitivity to the sound units of language, such as words, syllables, morphemes, and phonemes, and the ability to manipulate these units (Anthony & Francis, 2005; Gillon, 2004). Print concept as a conceptual literacy skill refers to a child's knowledge of the forms, conventions, and functions of print, such as understanding

the directionality of print, and that print conveys meaning (Adams, 1990; Justice & Ezell, 2001; Weiss & Hagen, 1988).

2.1.2 Concurrent Relationships among Literacy and Language Skills

Previous research has yielded robust findings on the concurrent correlations among oral language (e.g., vocabulary), literacy (e.g., print concept), and meta-linguistic skills (e.g., phonological awareness). From as early as age 3, these skills have been shown to have strong correlations (e.g., Bowey, 1994; Burgess & Lonigan, 1998; Dickinson et al., 2003; Dickinson & Snow, 1987; Lonigan et al., 1998; Storch & Whitehurst, 2002; Townsend & Konold, 2010), which also occur during kindergarten and early primary school (e.g., Duff et al., 2015; Lonigan et al., 2008; Storch & Whitehurst, 2002; Tabors et al., 2001; Wagner et al., 1993; Wagner et al., 1997; Wise et al., 2010). Furthermore, longitudinal studies have demonstrated that these concurrent correlations tend to sustain overtime, from preschool to primary school (e.g., Chaney, 1992, 1994, 1998; Hecht et al., 2000). Particularly, the lexical restructuring hypothesis explains the concurrent relationship between vocabulary and phonological awareness, by positing that phonological awareness increases as the size of vocabulary grows, to avoid potential confusion by lexical items with similar sounding (Goswami, 2001; Metsala, 1999). A large size of vocabulary is also likely to facilitate the development of print concept (Dickinson & Snow, 1987; Ramsook et al., 2020).

Crucially, the concurrent correlations between vocabulary, phonological awareness, and print concept are present across languages. The studies mentioned above were all based on children speaking English, an alphabetic language (Holm & Dodd, 1996), from countries including the US (e.g., Burgess & Lonigan, 1998; Dickinson et al., 2003; Wise et al., 2010), the UK (e.g., Duff et al., 2015; Hamilton et al., 2016), Canada (e.g., Sénéchal et al., 2001), and Australia (e.g., Bowey, 1994). In fact, these correlations have also been revealed among other monolinguals and even bilingual children. McBride-Chang et al. (2005) investigated the concurrent relationships between phonological awareness and vocabulary among second graders from Beijing, Hong Kong, Korea, and the US. Their findings highlighted that phonological awareness and vocabulary were closely correlated in English and some non-alphabetic languages; namely, Mandarin Chinese, Cantonese and Korean. Furthermore, these correlations are generalisable to monolinguals speaking other alphabetic languages, such as Dutch (Janssen et al., 2017) and French (Chung et al., 2021), bilinguals speaking two alphabetic languages, such as Spanish-English (Uchikoshi, 2019), French-English (Chung et

al., 2021), and Turkish-Dutch (Janssen et al., 2017), as well as those speaking one alphabetic language and one non-alphabetic language, such as Mandarin-English (Bialystok et al., 2003) and Cantonese-English (Uchikoshi, 2019). Overall, the concurrent correlations between early literacy and language skills, especially between vocabulary, phonological awareness, and print concept, seem strong cross-linguistically and -culturally.

2.1.3 Developmental Relationships among Literacy and Language Skills

Apart from the concurrent relationships, the roles of early literacy and language skills as developmental precursors to later literacy and language outcomes have been demonstrated by longitudinal studies (e.g., Hood et al., 2008; Storch & Whitehurst, 2002).

Later reading ability as a procedural literacy skill has long been associated with vocabulary, phonological awareness, and print concept (Pullen & Justice, 2003). The simple view of reading posits two components of reading, decoding and comprehension (Hoover & Gough, 1990). Decoding is the ability to transform printed letter strings into phonological units (Perfetti, 1985), which implies the role of phonological awareness as a prerequisite. Indeed, considerable evidence has shown that phonological awareness is key to the acquisition of decoding skills (Ehri et al., 2001; Rack et al., 1992; Torgesen et al., 1999). Reading comprehension refers to children's understanding of the print, at lexical, syntactic, semantic, and pragmatic levels (Aarnoutse et al., 2001; Hoover & Gough, 1990), which can be predicted by literacy and language skills like vocabulary and grammar (Muter et al., 2004). Reading ability overall is influenced by precursor skills, according to longitudinal studies. For instance, Storch and Whitehurst (2002) followed 626 children from preschool to later primary school and revealed that oral language and code-related skills (e.g., print concept, phonological awareness) played a critical role in children's reading ability throughout early primary school. Likewise, Sénéchal and LeFevre (2002) found that oral language and phonological awareness at school entry directly predicted children's reading at the end of Grade 3.

Given that literacy is multidimensional (Sénéchal et al., 2001), we intended to include more outcome measures other than reading. For example, writing is also considered essential for academic success (Graham & Harris, 2005). Considering the bidirectional relationships between reading and writing (Berninger et al., 2002), predictors of reading ability are expected to be associated with writing ability. However, research on predictors of primary

school writing is scarce and existing findings appear inconclusive. While the roles of vocabulary, letter and word reading at school entry in predicting writing during later primary school have been reinforced (e.g., Dunsmuir & Blatchford, 2004; Kim et al., 2015), the predictive role of phonological awareness has been supported in some (e.g., Silva & Martins, 2003) but not others (e.g., Hooper et al., 2010). Besides, print concept has been recognised as a building block of writing in kindergarten (Ritchey, 2008), yet little is known about its longitudinal relationship with writing ability in primary school.

Further, children's abilities to speak and listen are oral language skills that are central to their communication in school (Backlund, 1985). It is widely recognised that speaking, listening, reading, and writing skills are stably and closely related during primary school (Berninger & Abbott, 2010). However, speaking and listening skills have only been examined in a few developmental studies as outcome measures. Kendeou et al. (2009) investigated cross-sectional cohorts and found that preschool skills (e.g., vocabulary, phonological awareness) predicted children's oral expression and listening abilities in kindergarten, which in turn predicted oral expression and listening abilities in early primary school. The fundamental role of vocabulary in the development of speaking and listening skills has also been reported by some (e.g., Florit et al., 2009; Muter et al., 2004), yet the roles of phonological awareness and print concept are relatively under-researched. Accordingly, we aimed to replicate some of the developmental relationships among literacy and language skills, and potentially reveal more longitudinal relationships using more outcome measures.

2.2 Family Background Predicting Early Literacy and Language Development

Given the importance of literacy and language development in school readiness and achievement, it is therefore vital to investigate the contributing factors of these skills to guide educational practices (Burchinal et al., 2002). At school entry, children bring prior experiences from home that are crucial determinants of literacy and language development during primary school (Storch & Whitehurst, 2001). Therefore, a considerable amount of research has explored the effect of family background on early literacy and language development (e.g., Hamilton et al., 2016; Pace et al., 2017; Suizzo & Stapleton, 2007).

2.2.1 Dimensions of Family Background

Family background encompasses a series of interrelated factors such as family composition (e.g., single parenthood), ethnicity, genes, and socioeconomic status (SES) (Ladd et al., 1999; Scarr & Weinberg, 1978), among which SES is a multidimensional construct including parental occupation, education, and family income (Ermisch & Francesconi, 2001; Suizzo & Stapleton, 2007). Research suggests that children from low-SES families tend to underperform on language assessments compared to their higher-SES counterparts (Hoff, 2013). According to Hart and Risley's (1995) influential research, children in poverty would have heard approximately 30 million fewer words than their affluent peers by age 4, meaning that SES disadvantages can impede the quantity of language inputs. Moreover, recent research demonstrates that SES-related differences in early language environments also manifest in the quality of input and interactions, other than quantity (Cartmill et al., 2013; Goldin-Meadow et al., 2014). These early disparities in language environments potentially create gaps in literacy and language outcomes that remain stable or broaden overtime (Farkas & Beron, 2004; Fernald et al., 2013; Walker et al., 1994), leaving children from low-SES backgrounds at risk of school under-attainment and longer-term adversity (Shonkoff et al., 2012). Collectively, the impact of SES as a factor of family background on literacy and language development is well documented.

However, one may question whether all SES indices predict literacy and language development equally well. Indeed, maternal education has been identified as the most powerful mechanism underlying SES, in predicting children's literacy and language skills (Bornstein et al., 2003). This has been supported by studies in high-, middle-, and low-income countries (Bornstein et al., 2003; Jeong et al., 2017). Several reasons may account for the centrality of maternal education. First, it is more stable overtime, than parental occupation and family income which can easily fluctuate (Huston et al., 1994). Second, a higher educational level brings increased knowledge and cognitive skills (e.g., problem-solving, literacy and language) to facilitate high-quality parenting (Suizzo & Stapleton, 2007). Third, education may promote a mother's self-efficacy in parenting, which reflects their belief in the effectiveness of their parenting and potentially facilitates the establishment of better language environments (Dauber & Epstein, 1993; Machida et al., 2002).

Therefore, this study employed measures of maternal education and family income around childbirth to capture the relationships between SES and early literacy and language

development, considering the significance of maternal education (Bornstein et al., 2003), and the seemingly detrimental effect of living in poverty (Hart & Risley, 1995).

2.2.2 Maternal Literacy and Numeracy Skills

Nonetheless, we argue that SES is not the only factor of family background contributing to children's literacy and language development. Crucially, twin studies indicate that children's literacy and language skills are also influenced by genetic factors, with a substantial heritability of approximately 70% (Christopher et al., 2015; de Zeeuw et al., 2015). Therefore, the longitudinal relationships between family background and children's literacy and language skills may well be explained by the genes passed on from parents to offspring (Christopher et al., 2015).

Indeed, recent studies have demonstrated that parental skills can extensively predict children's skills. Puglisi et al. (2017) showed that maternal language and phonological skills can independently predict children's oral language and procedural literacy skills (e.g., reading). Similarly, van Bergen et al. (2017) found that both paternal and maternal procedural literacy, especially decoding skills can independently explain a large variance in children's decoding skills. Further, in both studies, other variables (e.g., storybook exposure) became non-significant predictors of children's literacy and language skills after parental skills were included in analysis (Puglisi et al., 2017; van Bergen et al., 2017), emphasising the unique role of parental literacy and language skills.

Correspondingly, the present study included measures on maternal skills around childbirth to provide a more comprehensive view of predictors of children's literacy and language, together with maternal education and family income. However, as a secondary analysis of the Millennium Cohort Study (MCS), our selection of variables was subject to data availability. The original team had to finish collecting data associated with the cohort members from multiple informants within strict timeframes, so performance data of children's parents concurrently in each sweep could not be collected (Hansen, 2014). Hence, instead of performance scores, we extracted mothers' self-reported difficulty in performing some daily activities to reflect their literacy and numeracy skills. The numeracy skills were also included because numbers are pronounced with verbal codes and are intertwined with literacy and language skills (Gal, 2002; Purpura & Napoli, 2015). Although little research has been conducted using parental self-reported skills to predict children's skills, we

hypothesised they would function similarly as assessment scores in other studies (e.g., Puglisi et al., 2017).

2.3 HLE Predicting Early Literacy and Language Development

Home literacy environment (HLE) is thought to be more proximal to children's literacy and language development than family background factors such as SES, although they are highly associated (Hamilton et al., 2016; Inoue et al., 2018). In fact, many studies have reported that the effect of family background, at least SES on early literacy and language skills is mediated by HLE (e.g., Aram et al., 2013; Forget-Dubois et al., 2009; Foster et al., 2005). Namely, HLE can predict children's literacy and language development over and above family SES. For example, Christian et al. (1998) highlighted that regardless of mother's educational level, kindergarten children who experienced a fertile HLE outperformed their peers with low HLE across oral language and literacy tasks. This stressed the importance of considering the role of HLE in this dissertation.

2.3.1 The Construct of HLE

HLE is a multidimensional construct, which usually refers to attributes of the home that are related to children's literacy and language learning (Burgess et al., 2002; Hamilton et al., 2016; Nag et al., 2019). One of the prominent theoretical frameworks on HLE is the home literacy model proposed by Sénéchal and LeFevre (2002). This model suggests that HLE can be classified into formal and informal interactions, which work independently to facilitate literacy and language development. Formal interactions are code-focused where parents engage children in the print by directly providing instructions, such as teaching children how to read alphabets. Informal interactions are meaning-focused where children are incidentally exposed to print through activities such as parent-child shared storybook reading (Liu et al., 2018; Sénéchal & LeFevre, 2002). In Sénéchal and LeFevre's (2002) original study, they measured informal interactions through storybook exposure, as indicated by parents' performance on two checklists: titles and authors of children's storybooks; whereas the formal interactions were measured via parent-reported frequencies of teaching their child to read and to print words.

Sénéchal and LeFevre's (2002) home literacy model and similar measurements have been employed by many studies examining whether HLE predicts literacy and language. Although this model was developed with a Canadian middle- and upper middle-class sample,

later studies have confirmed its generalisability to other countries, such as Australia (e.g., Hood et al., 2008), China (e.g., Liu et al., 2018), and the UK (e.g., Puglisi et al., 2017), to low- and middle-SES contexts (e.g., Hood et al., 2008), as well as to children at family-risk of dyslexia (e.g., Hamilton et al., 2016). Hence, the home literacy model appears valid across various contexts and populations.

2.3.2 Extending Beyond the Home Literacy Model

However, we argue that there are more attributes of home that should be measured in the operationalisation of HLE. Indeed, Sénéchal and LeFevre (2002) initially collected parent-report data on more aspects of HLE, including the frequencies of shared reading, library visits, and the child initiating shared reading, and the number of children's books at home, which were eventually excluded from analysis as they were not predictive of child outcomes. Nonetheless, Hood et al. (2008) found that in their sample, some of these variables added to the prediction of literacy and language skills. This was also consistent with Frijters et al.'s (2000) findings and the idea that multiple measurement items may account for more variance in children's literacy and language outcomes. Therefore, we also intended to include more HLE measures (e.g., the frequencies of shared reading and library visit) and test their predictive power in our sample.

Moreover, the operationalisation of HLE in previous studies was limited to parent-child interactions involving conventional literacy skills, namely, reading and writing (Hood et al., 2008; Sénéchal & LeFevre, 2002). However, since HLE is defined as home attributes related to literacy and language learning (Burgess et al., 2002; Nag et al., 2019), interactions in other domains may also benefit such development. For instance, several studies have emphasised the supportive roles of art- (e.g., Heath & Wolf, 2005), music- (e.g., Paquette & Rieg, 2008) and numeracy-related (e.g., Purpura & Napoli, 2015) activities in literacy and language development. Consequently, like Nag et al. (2019), we adopted a broader scope in conceptualising HLE, by examining more attributes of home at age 3 that can potentially foster literacy and language development.

Regarding formal interactions, apart from teaching on reading and writing, parental instruction in other domains may also facilitate literacy and language development, such as instruction on numbers and songs. The reciprocal relationships between numeracy and literacy/language domains are well established. For example, Duncan et al.'s (2007) analysis

on multiple US and UK datasets indicated that early numeracy skills at school entry accounted for more variance in later literacy outcomes, compared to early literacy skills. Meanwhile, Purpura and Napoli's (2015) study on a US sample revealed that vocabulary can predict early numeracy skills (e.g., counting), which in turn predicted later print concept. Hence, numeracy instruction from parents may additionally facilitate literacy and language skills, alongside their numeracy skills (Dulay et al., 2019). Similarly, music, rhythm and rhyme can enhance children's literacy and language abilities, which seems applicable to both first language (Hill-Clarke & Robinson, 2004) and second language learning (e.g., Gan & Chong, 1998; Paquette & Rieg, 2008). Particularly, children's skills such as phonological awareness, listening comprehension and articulation tend to be heightened during musical interactions with parents (O'Herron & Siebenaler, 2007). Lamb and Gregory (1993) reported that children who were good at discriminating pitch also had high phonological awareness and reading skills, aligned with findings from musical intervention studies (e.g., Gromko, 2005). Altogether, there is evidence for the association of numeracy and music instruction with early literacy and language development. Correspondingly, this study used the term 'home tutoring' (e.g., Nag et al., 2019, p. 91) instead of literacy instruction (e.g., Hamilton et al., 2016), to capture a diverse range of parental instruction relevant for children's literacy and language learning.

Furthermore, we used 'home literacy practices' to reflect more aspects of informal HLE, other than storybook exposure (e.g., Sénéchal & LeFevre, 2002). This term was altered from Nag et al.'s (2019, p. 91) term 'adult literacy practices', referring to parents' resources, values, and attitudes towards literacy education. These aspects of HLE can influence parents' choices of parent-child leisure-time activities and how to dispense resources (Burgess, 2005). For example, a parent who values literacy and language learning may invest more on children's enrichment resources, spend more time reading to the child or increase the frequency of library visit to expand the home collection of children's books (Burgess et al., 2002). Similarly, rather than leaving the child to watch alone, they may engage in parent-child interactions during television viewing (Liebeskind et al., 2013), which has been found beneficial for literacy and language development (Reiser et al., 1984). Instead of using 'adult literacy practices' (Nag et al., 2019, p. 91), which may imply that it is mainly the parents who are performing these activities, we wanted to emphasise the participation of both parents and children, by using the term 'home literacy practices'.

2.3.3 Relationships between Family Background and HLE

As mentioned previously, HLE is likely to mediate the relationships between family background, at least SES and children's literacy and language skills (e.g., Aram et al., 2013; Forget‐Dubois et al., 2009), which also suggests an association between family background and HLE. Indeed, how family income, maternal education, and skills each are linked with home tutoring and literacy practices has not been discussed. Despite the trend that families with more-educated mothers are likely to have higher income, the relationship between family income and maternal education is not monotonic. Even in low-income families, mothers with some educational qualifications may provide an adequate HLE for their children (Hartas, 2011). Therefore, it is crucial to distinguish between these two variables when investigating how SES and HLE are linked.

Findings regarding the associations between the two SES indices and home literacy practices appear to be consistent across countries. Generally, families with higher income and more-educated mothers are more likely to provide home literacy practices, through activities such as shared storybook reading, as well as more access to literacy resources such as children's books. This has been revealed by analyses on national longitudinal surveys in the US (e.g., Carneiro et al., 2013) and the UK (e.g., Hartas, 2011), and on smaller samples from some other high-income countries (Canada: Sénéchal, 2006; Netherlands: Van Steensel, 2006; Europe: Hemmerechts et al., 2017). Furthermore, studies on low- and middle-income countries have yielded similar findings. Kalia and Reese (2009) reported that in a sample of Indian children, maternal education and household income were positively associated with book-reading practices at home. Farver et al.'s (2006) research on a sample of Spanish-English bilingual children from low-income Latino families in the US also confirmed this trend. Overall, family income and maternal education are predictive of home literacy practices, especially book reading.

Contrarily, fewer studies have been conducted on home tutoring and the picture on the relationships between SES and home tutoring is more complex. Sénéchal (2006) followed a Canadian sample and showed that parental education was positively related with home tutoring on alphabets, reading, and writing during kindergarten. However, Hamilton et al. (2016) found family SES, indexed by parental education and occupation, not predictive of home tutoring on the same areas in a UK sample. Similarly, studies in developing countries, such as China (e.g., Liu et al., 2018) and India (e.g., Kalia & Reese, 2009), suggested non-

significant associations between family SES, indexed by parental education and family income, and home tutoring on reading and writing. Moreover, Hartas's (2011) analysis on the MCS dataset indicated that maternal education and family income can predict home tutoring on songs/poems/rhymes but not that on alphabets. Accordingly, most studies suggest that parents from diverse SES backgrounds tend to offer home tutoring equally frequently (Hartas, 2011), except for some inconsistent findings (e.g., Sénéchal, 2006).

Further, little is known about how mother's literacy and numeracy skills in daily activities are associated with HLE. Previous studies drew conclusions based on performance data, which the current study could not achieve as a secondary data analysis. For example, van Bergen et al. (2017) assessed parents' and children's reading fluency, as well as HLE (parents' reading frequency, home subscriptions to magazines/newspapers, number of books at home), reporting significant correlations between maternal reading fluency all three HLE variables. However, van Bergen et al.'s (2017) operationalisation of HLE was distinct from ours, as they did not measure parent-child interactions. Correspondingly, Puglisi et al.'s (2017) research may be more relevant for our study, as they measured home tutoring (on letters, reading, writing) and home literacy practices (e.g., number of children's storybooks at home, frequency of shared reading, parental familiarity with the title and authors of children's books). Meanwhile, maternal language and phonology were formally tested using cognitive assessments. They found maternal language predictive of home literacy practices but not home tutoring, whereas maternal phonology was not predictive of HLE overall. Collectively, there is some evidence that mother's literacy and language skills can predict some attributes of home literacy practices (Puglisi et al., 2017; Sénéchal et al., 2001), which may apply to relationships between daily literacy/numeracy skills and informal HLE. Nevertheless, the scarcity of studies involving maternal self-reported skills warrants further research.

2.4 Predicting Literacy and Language Skills with HLE and Family Background

Having discussed the concurrent and longitudinal relationships among early literacy and language skills, and the relationships between family background and HLE, it is now vital to review how family background and HLE predict children's skills. Globally, HLE tends to be significantly and longitudinally associated with children's reading abilities. Analyses based on the Progress in International Reading and Literacy Study (PIRLS) dataset, such as Arya et al.'s (2014) on 52 countries and Park's (2008) on 25 countries, have

demonstrated that home literacy practices (e.g., read books, sing songs, play with alphabet toys) are predictive of children's reading abilities, even after controlling for family SES. However, they also highlighted the possible heterogeneity in this relationship across countries, due to different levels of economic development. Therefore, a review of this series of relationships within individual countries is needed.

Several lines of research have investigated whether home literacy practices can predict children's oral language, literacy, and meta-linguistic skills longitudinally, producing mixed results. Generally, the role of home literacy practices in predicting oral language skills (e.g., vocabulary, listening comprehension) has been supported across countries and several language systems (Australia: Hood et al., 2018; Canada: Sénéchal & LeFevre, 2002; China: Liu et al., 2018; Israel: Korat et al., 2013; Netherlands: Van Steensel, 2006; UK: Hamilton et al., 2016). In terms of literacy skills, some studies (e.g., Hamilton et al., 2016; Korat et al., 2013; Van Steensel, 2006) found home literacy practices predictive, but not others (e.g., Hood et al., 2008; Liu et al., 2018; Sénéchal & LeFevre, 2002). This may be due to the difference in measurements of literacy skills. For example, Sénéchal and LeFevre (2002) developed a composite score of children's results from print concept, alphabet knowledge, decoding and invented spelling tests, whereas Hamilton et al.'s (2016) composite literacy score was based on reading and word recognition tests. Meanwhile, there are also differences in the measurement of home literacy practices, as some only focused on storybook exposure (e.g., Hamilton et al., 2016; Sénéchal & LeFevre, 2002), whilst others examined more aspects such as library visit (Hood et al., 2018; Van Steensel, 2006), educational television viewing (Van Steensel, 2006), and educational games at home (Korat et al., 2013). Possibly because of these measurement differences, findings on the association between home literacy practices and phonological awareness are also inconsistent, with some significant (e.g., Korat et al., 2013; Van Steensel, 2006) and some non-significant results (e.g., Hamilton et al., 2016; Hood et al., 2008; Liu et al., 2018; Sénéchal & LeFevre, 2002). Collectively, home literacy practices are usually predictive of children's oral language skills, but its role in literacy and meta-linguistic skills is inconclusive.

Unlike home literacy practices, the picture on the role of home tutoring is relatively less complex. Specifically, the home literacy model posits that home tutoring can predict code-related literacy skills, such as print concept, decoding, alphabet knowledge, letter knowledge, and word knowledge (Martini & Sénéchal, 2012; Sénéchal & LeFevre, 2002),

which has been supported by research across countries, including Australia (Hood et al., 2008), Canada (Inoue et al., 2018; Stephenson et al., 2008), and the UK (Hamilton et al., 2016). Meanwhile, home tutoring is not predictive of oral language skills as they are not code-related (Martini & Sénéchal, 2012), and no study in this synthesis of relevant literature has revealed a significant association between these two constructs. However, these studies yielded different findings considering the relationship between home tutoring and phonological awareness, as some found home tutoring predictive (e.g., Hamilton et al., 2016; Inoue et al., 2018; Liu et al., 2018), but not others (e.g., Hood et al., 2008; Sénéchal & LeFevre, 2002; Stephenson et al., 2008). Accordingly, home tutoring appears predictive of early literacy skills but not oral language, whereas its association with phonological awareness is uncertain.

Additionally, several mediated relationships should be noted when analysing these developmental relationships. HLE has been reported to mediate the relationship between family SES and children's literacy and language skills (e.g., Aram et al., 2013; Forget-Dubois et al., 2009; Foster et al., 2005). Moreover, previous research has indicated that HLE can indirectly predict children's later literacy and language outcomes via the mediation of precursor skills (e.g., De Jong & Leseman, 2001; Hamilton et al., 2016; Inoue et al., 2018). Hence, apart from the direct relationships, we aimed to replicate some of these indirect relationships, by linking family background around childbirth with both early and later literacy and language skills, and by linking HLE variables with later outcomes.

2.5 Theoretical Model

Hamilton et al.'s (2016) model (Figure 1) was adopted to guide most predictions tested in this study. This is because we intended to examine similar direct and indirect developmental relationships presented in their model. Like Hamilton et al. (2016), the current study included four sets of variables, family background, HLE, early skills, and later outcomes, following children from preschool to early primary school. Although the family background variables in our study were assessed around childbirth, in contrast to theirs at age 4, we expected the effects of these variables to be stable, especially maternal education and literacy/numeracy skills (e.g., Inoue et al., 2018).

Figure 1

Longitudinal Path Model Predicting Word-level Literacy and Reading Comprehension

Note. Sample was children with no family-risk of dyslexia. SES = socioeconomic status. Dashed lines represent nonsignificant pathways. Adapted from "The Home Literacy Environment as a Predictor of the Early Literacy Development of Children at Family-Risk of Dyslexia," by L. G. Hamilton, M. E. Hayiou-Thomas, C. Hulme, and M. J. Snowling, 2016, *Scientific Studies of Reading, 20*(5), p. 412 (https://doi.org/10.1080/10888438.2016.1213266). Copyright 2016 by the Taylor & Francis Group.

2.5.1 Methodological Advantages of Hamilton et al.'s Model

Apart from similar research interests, the methodological advantages of Hamilton et al.'s study (2016) also facilitated the decision to replicate their model. A major advantage was related to their use of an advanced statistical technique – path model, a type of structural equation modelling (SEM). Earlier studies on this topic relied heavily on regression analyses (e.g., Hartas, 2011; Hood et al., 2008; Sénéchal & LeFevre, 2002; Van Steensel, 2006). Compared to regression models which estimate the relationship between predictors and only one outcome variable at a time, SEM allows multiple related equations to be calculated

simultaneously (Hoyle, 1995; Quinn & Wagner, 2018). Meanwhile, SEM explicitly specifies the measurement errors that seem inevitable in social sciences, providing a more realistic view of the measures than regression which tends to assume that measurement occurs without error (Kline, 2015, p. 13). Accordingly, SEM was also employed in this study.

2.5.2 Going Beyond Hamilton et al.'s Model

We intended to extend Hamilton et al.'s (2016) model in several ways. First, we attempted to replicate some of their predictions in a national population from Wales, with data obtained from the MCS, whereas Hamilton et al. (2016) focused on comparing the literacy development between children at family-risk of dyslexia and those with no known risk, with a relatively small sample size. Second, we considered both the environmental (i.e., maternal education, family income) and genetic (i.e., maternal literacy/numeracy skills) components of family background and linked them individually with HLE, instead of jointly, as they may have different relationships with home tutoring and home literacy practices (Hartas, 2011; Puglisi et al., 2017). Third, we used different terms for formal and informal HLE, and included more measures in both categories other than those related to reading and writing. Specifically, formal HLE was termed 'home tutoring,' including parental instruction on alphabets, numbers/counting, and songs/poems/rhymes. Informal HLE was named 'home literacy practices,' including the frequencies of parent-child book reading, library visit, watching television/videos, and painting/drawing (after Nag et al., 2019). Finally, we employed a mixture of child performance data and teacher-report ratings as measures of children's literacy and language skills, subject to data availability. Particularly, we obtained teacher ratings on children's phonological awareness and print concept at age 5, and their English speaking/listening, reading, and writing at age 7.

2.6 The Present Study

We examined the developmental relationships between family background around childbirth, HLE at age 3, early, and later literacy and language skills from ages 5 to 7. Using the Wales subsample from the MCS, we aimed to address three main questions listed in Section 1.4.

The hypothesised model is shown in Figure 2. In the Wales sample, we expected family background to be related to HLE, and to literacy and language skills at ages 5 and 7 via full mediations. Specifically, we expected family income and maternal education to

predict home literacy practices, but not home tutoring (Hartas, 2011; Puglisi et al., 2017). Maternal skills were also linked with both aspects of HLE, but no specific predictions were made about these relationships, due to the lack of previous research involving maternal self-reported literacy/numeracy skills. Moreover, we expected home tutoring to predict phonological awareness and print concept (Hamilton et al., 2016; Inoue et al., 2018), while home literacy practices to predict all three early skills (Hartas, 2011; Korat et al., 2013; Van Steensel, 2006). HLE would also predict later outcomes at age 7 via the full mediation of precursor skills (De Jong & Leseman, 2001; Inoue et al., 2018). Furthermore, we expected each precursor skill at age 5 to predict each literacy and language outcome at age 7. Concurrent correlations among literacy and language skills were also predicted (Berninger & Abbott, 2010; Kim et al., 2015).

Figure 2

Hypothesised Structural Equation Model Predicting Literacy and Language Skills at Age 7

Note. Single-headed arrows represent pathways between variables. Double-headed arrows represent correlations between variables. To simplify presentation, direct pathways in hypothesised mediated relationships are not shown.

Chapter 3: Methodology

This chapter will outline and justify the methodology incorporated in this study. First, the database from which the data was drawn will be introduced, followed by an explanation of the analytic sample and an evaluation of the secondary data analysis approach. Next, ethical considerations will be examined. This chapter will then discuss the measures examined in this study. Finally, the data analysis process will be explicated, including the statistical model and the method to handle missing data.

3.1 Database

The data for this study was drawn from the Millennium Cohort Study (MCS), a national longitudinal birth cohort study in the UK funded by the Economic and Social Research Council (ESRC). It was designed to capture the development and family circumstances of around 19,000 children born at the beginning the new century, across England, Scotland, Wales, and Northern Ireland. The MCS measures the cohort members' physical, socio-emotional, cognitive, and behavioural development via age-appropriate assessments. Reports on their daily life, behaviour and experiences are also gathered from the children themselves, resident parents, sometimes teachers and older siblings. Additionally, data on the children's living context is collected, including their families' socio-economic status, parenting, childcare, and relationships (Hansen, 2014).

3.2 The Sample

3.2.1 Sample Design

The MCS sample was drawn from a population of children born between 1 September 2000 and 31 August 2001 for England and Wales, and between 24 November 2000 and 11 January 2002 for Scotland and Northern Ireland, who were alive and living in the UK at 9 months of age, whose families were eligible to receive Child Benefit at that time (Plewis, 2007a, p. 7). Eligible children were identified via the Child Benefit register, a benefit with virtually universal coverage except those ineligible due to recent or temporary immigrant status (Connelly & Platt, 2014).

To ensure representativeness overtime, the sample was designed as a disproportionately stratified cluster sample. Specifically, the sample was clustered geographically using electoral wards as a sampling frame, which are small areas used to

organise local government elections in the UK (Plewis, 2007b). Meanwhile, the sample was stratified to allow for over-representation of families with high levels of poverty based on the Child Poverty Index, and those from areas with relatively high ethnic minority concentration (Hansen, 2014). In England, the population was stratified into three strata – advantaged, disadvantaged, and ethnic minority. Whereas in Wales, Scotland and Northern Ireland, there were only two strata, advantaged and disadvantaged, considering the low percentages of ethnic minority groups (approximately 1% of the population) (Plewis, 2007a, p. 9).

3.2.2 Recruitment

The target sample was randomly selected from each stratum of each country and contacted by the Department for Work and Pensions for recruitment (Plewis, 2007a, p. 20). The first sweep achieved a sample of 18,818 cohort members, including some twins and triplets, from 18,552 families clustered in 398 electoral wards (Hansen, 2014). For the subsequent sweeps, the participating families at the previous sweep would be contacted again for participation, if they remained living in the UK at the time of sampling (Plewis, 2007a, p. 7). To date, there have been seven surveys, following the cohort members and their families from years 2001-2002 at the age of 9 months, to the later sweeps at the ages of 3, 5, 7, 11, 14, and 17 years. The most recent sweep during years 2018-2019 when the cohort members reached the age of 17, achieved a sample of 10,757 cohort members from 10,625 families (Fitzsimons et al., 2020).

3.2.3 The Analytic Sample

The analytic sample was selected according to our research focus. First, only singleton cohort members and one child from each twin and triplet births were included, to avoid duplicate family- and parent-level data. Second, only families living in Wales throughout sweeps 1 to 4 were included, considering the research focus on Welsh children. Finally, cohort members with high levels of missingness on literacy and language variables at sweeps 3 and 4 were omitted. Consequently, the analytic sample contains 612 cohort members resident in Wales from sweep 1 to 4 (315 males, 297 females).

The parent-report measures were completed by primary caregiver as the main respondent, with some parallel information collected from a partner respondent. At sweep 1, the main respondent was the biological mother in all cases. In 610 cases (99.7% of the analytic sample), the partner respondent was the biological father; and in two cases (0.3%),

the stepfather or partner of the biological mother. This profile changed by sweep 2 (Appendix A), due to some changes in family composition.

3.3 Secondary Data Analysis

This study utilised the existing MCS database to answer new research questions and was therefore a secondary data analysis – an analysis of data initially collected and analysed for another primary purpose (Dale et al., 2008; Kielcolt & Nathan, 1986).

A major advantage of this approach is its cost-effectiveness in terms of time, finance, and personnel (Dale et al., 1988; Hyman, 1972). Indeed, large-scale data with such high quality and breadth of variables as the MCS would not have been collected within the timeframe of this dissertation. Moreover, quantitative secondary datasets tend to have used validated measurements and recruited large samples that are designed to be representative of the target population (Dale et al., 2008; Smith et al., 2011). A representative sample is advantageous as it facilitates the generalisation of research findings to wider populations (Dale et al., 2008). Additionally, secondary data analysis enables triangulation with data from other sources (Smith, 2008). Here, it allows us to compare results from a longitudinal national survey with those from Hamilton et al.'s (2016) smaller-scale longitudinal study on a predetermined child population.

Despite the advantages, the absence of the secondary analyst during the original data collection procedure is viewed as a methodological issue, as they may not fully understand how the data were constructed (Dale et al., 2008; Johnston, 2017). This may lead to misinterpretation of data and constructs, which is problematic (Kielcolt & Nathan, 1986). Therefore, we familiarised with the data via other means such as technical reports, codebooks, and manuals, and attempted to recognise the strengths and limitations of the dataset prior to analysis (Dale et al., 1988; Kielcolt & Nathan, 1986).

Another relevant concern is whether the data can answer the novel research questions proposed by the secondary analyst (Dale et al., 2008; Johnston, 2017). The data was collected for another primary purpose and may lack the information that the secondary analyst would like to collect (Boslaugh, 2007; Dale et al., 2008). We, however, believe that the MCS can adequately address our research questions, although some measures may not be optimal subject to data availability, which will be further explicated in the Discussion chapter.

3.4 Ethics

The MCS team received ethical approval from ethical committees before each sweep of surveying (see Hansen, 2014 for full details). Written informed consent was obtained from all adult respondents (e.g., parents) and from a parent or guardian for the participation of a cohort member under age 16 (Hansen, 2014).

The MCS data can be obtained from the UK Data Archive at the University of Essex. After user and project registration with the UK Data Archive, access to the MCS data was approved, and the datasets from sweeps 1-4 were downloaded onto an encrypted personal computer in anonymised SPSS format. As the downloaded data were publicly available and fully anonymised with no personally identifying information, no further ethical approval was sought for the current study (Dale et al., 2008).

3.5 Measures and Procedure

The measures involved in this study included: (a) family background at sweep 1 (when the cohort members were 9 months old), (b) the home literacy environment (HLE) at sweep 2 (3 years), (c) early literacy and language skills at sweep 3 (5 years), and (d) later literacy and language outcomes at sweep 4 (7 years). Family background and HLE measures were based on parental interviews, whereas the literacy and language measures were obtained from a combination of child cognitive assessments and teacher reports.

3.5.1 Family Background (Age 9 Months)

Maternal Education. The highest educational qualification attained by the mother by sweep 1 was measured on a six-point scale, equivalent to the National Vocational Qualification (NVQ) scale: 0 = no formal qualifications, 1 = General Certificate of Secondary Education (GCSE, exams taken at the end of compulsory education at age 16 in the UK) grades D-G, 2 = GCSE grades A*-C, 3 = A levels (exams taken at the end of secondary education around age 18 in the UK), 4 = diplomas in higher education, 5 = higher degree or first degree qualifications (National Council for Vocational Qualifications, 1995).

Family Income. Family income data were adjusted for the number of people in household using the equivalised income scale created by the Organisation for Economic Co-operation and Development (OECD), to generate a dichotomous indicator for relative poverty (0 = below 60% median, 1 = above 60% median). Families with disposable and market

income below 60% of the national median income are considered as living in relative poverty (Piacentini, 2014).

Mother's Literacy/numeracy Skills. Three aspects of mother's literacy/numeracy skills were self-reported: abilities to read aloud a children's storybook, to read and fill out forms, and to calculate change during shopping. Specifically, the mothers were asked in interviews whether they can usually perform each task easily or with difficulty, with three possible answers: 'Yes, easily', 'Yes, with difficulty', and 'No', meaning the task was not usually performed. The 'Yes, with difficulty' and 'No' responses were subsequently combined to form dichotomous variables on mother's literacy/numeracy skills as '1 = high' or '0 = low', which were used in the analysis instead.

3.5.2 Home Learning Environment (Age 3 Years)

Home Tutoring. Three attributes were assessed: teaching on alphabets, numbers/counting, and songs/poems/rhymes. At sweep 2, the interview items asked the main respondents to report whether and how often anyone in the household helped the cohort members to learn each of the three aspects. The frequency of home tutoring on each aspect was estimated on an 8-point scale: 0 = not at all, 1 = occasionally or less than once a week, 2 = 1-2 days per week, 3 = 3 times a week, 4 = 4 times a week, 5 = 5 times a week, 6 = 6 times a week, and 7 = 7 times a week.

Home Literacy Practices. This was measured by four items from the parent interviews – frequencies of the child being read to, visiting libraries, watching television/videos, painting/drawing.

Being Read to. The frequency of the cohort member being read to was measured by asking the main and partner respondent to estimate how often they themselves and anyone else read to the child at home on a 6-point scale: 1 = not at all, 2 = less often, 3 = once or twice a month, 4 = once or twice a week, 5 = several times a week, 6 = every day. The highest score among the ratings for the main respondent, partner respondent and anyone else at home was extracted for each family and used as a composite score of 'being read to' in analysis.

Library Visit. The main respondent was asked 'Does anyone at home ever take the child to the library?' and 'How often does someone at home take the child to the library?'.

The answers ‘No’ to the first question were treated as ‘not at all’ to the second question. Therefore, the frequency of library visit was measured on a 5-point scale: 0 = not at all, 1 = on special occasions, 2 = once a month, 3 = once a fortnight, 4 = once a week.

Watching Television/videos. The frequency of the cohort member’s television/videos watching was obtained from the main respondent interview, via their responses to ‘Typically, how many hours a day does the child watch television or videos?’. Rating was given on a 4-point scale: 1 = not at all, 2 = up to one hour, 3 = more than 1 hour, less than 3 hours, 4 = more than 3 hours.

Painting/drawing. The main respondents were also asked ‘Does the child ever paint or draw at home?’ and ‘How often does the child paint or draw at home?’. If the answer ‘No’ was given to the first question, the answer to the second question would be marked as ‘not at all’. Overall, the frequency of the child painting/drawing at home was rated on an 8-point scale: 0 = not at all, 1 = occasionally or less than once a week, 2 = 1-2 days per week, 3 = 3 times a week, 4 = 4 times a week, 5 = 5 times a week, 6 = 6 times a week, 7 = 7 times a week.

3.5.3 Early Literacy and Language Skills (Age 5 Years)

Vocabulary. Vocabulary at age 5 was assessed using a subtest from the British Ability Scale II (BAS II), the BAS Naming Vocabulary (Elliott, Smith, & McCulloch, 1996). In this assessment, the cohort member was shown one page from a booklet of 36 coloured pictures of objects at a time (e.g., ear, hourglass), and asked to name the object verbally in English. The raw score (i.e., the number of correct responses) of each child was transformed into a standardised score which had been adjusted for item difficulty and age. The standardised scores ranged from 20 to 80, with a mean of 50 and a standard deviation of 10 (Connelly, 2013; Elliott et al., 1996).

Phonological Awareness. Estimation on cohort member’s phonological awareness was derived from one subscale of teacher’s questionnaire in Wales, which replicated the Foundation Stage Profile (FSP) recording students’ achievement reported by teachers in England (Johnson, 2008; Qualifications and Curriculum Authority, 2000). Within teacher-report on ‘Communication, Language and Literacy (CLL)’, the subscale ‘Linking Sounds and Letters (LSL)’ was identified as an index for phonological awareness. The LSL scale consisted of nine items, each of them representing an attribute of phonological awareness, with an example item being ‘(The child) shows an awareness of rhyme and alliteration’.

Teachers were asked whether the child attained the attributes by assigning '0 = No' or '1 = Yes' to each item. A raw LSL score was calculated for each cohort member by summing teacher's ratings for nine items, ranging from 0 to 9 (Gray & Simmonds, 2009; Johnson, 2008).

Print concept. In teacher's questionnaire, the subscale 'Reading' from the CLL scale provided information on cohort member's print concept. Again, nine items were included in this subscale (e.g., '(the child) knows that in English, Irish or Welsh, print is read from left to right and top to bottom'), and the teachers were asked to choose from '0 = No' and '1 = Yes' for each item. Raw reading scores ranged from 0 to 9 (Gray & Simmonds, 2009; Johnson, 2008).

3.5.4 Later Literacy and Language Outcomes (Age 7 Years)

Cognitive Assessment. Children's ability to decode printed English words was assessed using the BAS Word Reading subtest (Elliott et al., 1996), where the cohort members were asked to read aloud a series of words presented on a card. This test consisted of 90 words in total, organised into nine blocks of ten words each, in an order of increasing difficulty (Hansen, 2014, p. 65). The raw scores (i.e., number of words read correctly) ranged from 0 to 90 and were adjusted for item difficulty and age to generate the standardised scores, which allowed comparisons with other assessment scores. Standardised scores ranged from 55 to 145, with a mean of 100 and a standard deviation of 15 (Connelly, 2013; Elliott et al., 1996), and were included in the analysis as 'English decoding'.

Teacher-report. Teachers filled out questionnaires to rate the cohort member's English speaking/listening, reading, and writing abilities respectively, in relation to all children around age 7. The three abilities were rated on an identical 5-point scale: 1 = well below average, 2 = below average, 3 = average, 4 = above average, 5 = well above average; and were included in the analysis separately.

3.5.5 Validity of Teacher Report

Unlike most previous studies using children's cognitive assessment outcomes as variables (e.g., Hamilton et al., 2016; Hood et al., 2008; Sénéchal & LeFevre, 2002), this study used a combination of assessment performance and teacher-report variables subject to data availability. A key question is then whether teacher judgement is a valid representation

of children's abilities. In fact, several studies have demonstrated that teacher ratings tend to strongly or moderately correlate with children's performance in terms of early literacy skills and academic attainment (e.g., Begeny & Buchanan, 2010; Beswick et al., 2005; Cabell et al., 2009; Meisels et al., 2001). Furthermore, Snowling et al. (2011) examined the validity of FSP in measuring children's current and future literacy and language skills and reported that teachers can provide accurate judgements about children's abilities when appropriately trained. Although teachers in Wales did not undertake training to fill out the FSP-equivalent teacher questionnaire in the MCS (Johnson, 2008), there may still be some degrees of validity in their judgements about children's literacy and language skills.

3.6 Data Analysis

The statistical analyses were conducted using SPSS 27 and Mplus 8.6 software. In the following subsections, more information about structural equation modelling (SEM), how it was applied to answering our research questions, goodness-of-fit indices, and the missingness and distribution of data will be provided.

3.6.1 Structural Equation Modelling (SEM)

SEM is a multivariate statistical technique used for representing and estimating models of linear relationships among observed (manifest) and unobserved (latent) variables (Hoyle, 1995; MacCallum & Austin, 2000). Latent variables correspond to hypothetical constructs presumed to represent a continuum that cannot be directly observed (Kline, 2015, p. 12). Each latent variable is generally represented by multiple observed variables that serve as indicators of this construct, and the latent construct can also be referred to as a factor (MacCallum & Austin, 2000). For example, the frequencies of parental instruction on three aspects (i.e., alphabets, numbers/counting, songs/poems/rhymes) may serve as three indicators for home tutoring as a latent construct. As mentioned, SEM is advantageous as it calculates multiple relationships simultaneously and explicitly specifies measurement errors that appear inevitable in social research (Hoyle, 1995; Kline, 2015, p. 13).

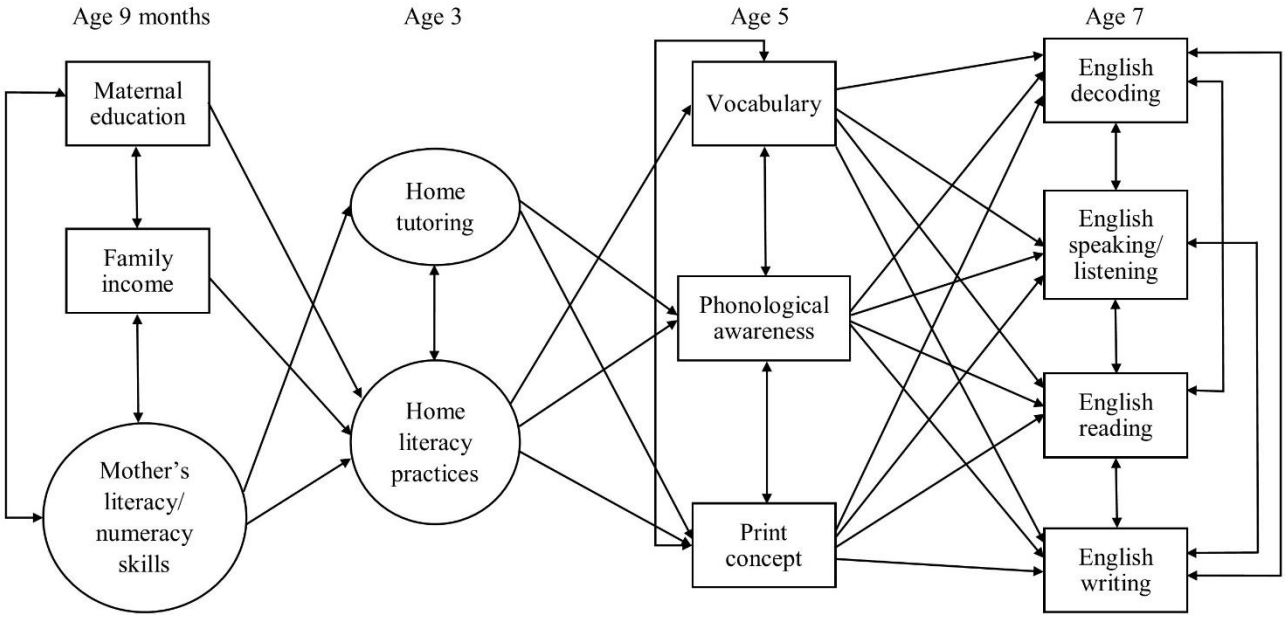
3.6.2 Hypothesised Model

A longitudinal structural equation model (Figure 3) was proposed for this dissertation. Instead of repeated measures, this study adopted different variables measured at successive sweeps of the MCS (sweeps 1-4) and attempted to specify a model to estimate the effects of variables at early sweeps on variables at later sweeps. The sequence and timing of

measurements at different sweeps of the MCS allow for the developmental relationships to be tested (MacCallum & Austin, 2000), as preschool family background and HLE preceded children's literacy and language development during school years, and the early skills at age 5 were precursors to the later outcomes at age 7.

Figure 3

Hypothesised Longitudinal Structural Equation Model Predicting Literacy and Language Outcomes at Age 7



Note. The latent constructs are in ovals and the observed variables are in rectangles.

This model was proposed as involving a combination of observed and latent constructs, where only mother's literacy/numeracy skills, home tutoring and home literacy practices were specified as latent constructs. The factorial validity of these latent constructs was examined using confirmatory factor analysis prior to conducting SEM (see Section 4.2). A confirmatory instead of exploratory factor analysis was conducted because there have been theoretical frameworks developed to classify HLE (e.g., Nag et al., 2019; Sénéchal & LeFevre, 2002), and performance on daily tasks can be viewed as having a common underlying factor (McGrew, 2009; Panizzon et al., 2014). Family background was operationalised as three separate constructs instead of one latent construct, because they were hypothesised to be associated with variables measured at later sweeps in different ways.

To ensure the representativeness of our results, the data were weighted to account for oversampling in sample design, non-response during sample recruitment at sweep 1, and attrition at later sweeps. This was implemented in Mplus by using the 'TYPE = COMPLEX' option of the 'ANALYSIS' command, together with the 'STRATIFICATION' and 'WEIGHT' options of the 'VARIABLE' command. The dichotomous variable of strata in Wales (0 = Wales-disadvantage, 1 = Wales-advantage) was used in the 'STRATIFICATION' command. The overall weight for single country analysis calculated by the MCS team was used in the 'WEIGHT' command, which had accounted for unequal selection probabilities of wards and adjusted for non-response (Hansen, 2014). Consequently, our analysis was likely to produce unbiased estimations for population parameters in the Wales context (Korn & Graubard, 1995).

3.6.3 Covariates

In our model, three covariates were controlled for to rule out their confounding influences on the dependent variables (Little, 2013, p. 16), as they were not the focus of our research questions.

Cohort members' sex and season of birth were controlled for, as males tend to underperform on language assessments compared to their female counterparts (e.g., Taylor et al., 2015; Taylor, 2018), and spring- and summer-born children are more likely to have lower attainment in primary school than those born in autumn or winter (e.g., Crawford et al., 2013; Taylor, 2018). Additionally, household single-parent status was controlled for, as being from a single-parent household has been suggested to affect children's educational outcomes (e.g., Taylor, 2018). By controlling for cohort members' sex, season of birth, and household single-parent status, the variance in literacy and language skills associated with these variables was explained away, so that the magnitude of effects of family background and HLE on literacy and language development can be estimated more accurately (Little, 2013, p. 15).

3.6.4 Goodness-of-fit Indices

An abundance of goodness-of-fit indices have been developed to indicate how well a structural equation model approximates the sample data (McDonald & Marsh, 1990; Schermelleh-Engel et al., 2003). In this dissertation, four goodness-of-fit indices were reported – the Chi-square statistic, the Comparative Fit Index (CFI), the Root Mean Square

Error of Approximation (RMSEA), and the Standardised Root Mean Square Residual (SRMR).

Chi-square Statistic. The chi-square statistic tests the null hypothesis that the sample data and hypothesised model do not differ (i.e., exact fit). Therefore, the hypothesised model is rejected if the statistic is significant (i.e., $p < .05$) (Bryne, 2009). However, this index has a major limitation of being sensitive to sample size. With large samples, the hypothesised model is often rejected, even if the model describes the data well (Hox & Bechger, 1998). Therefore, other indices were also used to indicate model fit, considering the increased possibility of inflated Type I error (i.e., incorrectly rejecting a model that fits well) with larger samples in chi-square statistic.

CFI. The CFI (Bentler, 1990) is less affected by the sample size compared to the chi-square statistic. Its value ranges from zero to one, with a higher value indicating a better fit. As a rule of thumb, values greater than .95 can be interpreted as an acceptable fit, whereas values greater than .97 are indicative of good fit (Hu & Bentler, 1999).

RMSEA. The RMSEA (Steiger, 1990) assesses whether the hypothesised model fits the data approximately well, which seems more sensible than the null hypothesis of exact fit (Browne & Cudeck, 1993). Browne and Cudeck (1993) claimed that RMSEA values ≤ .05 can be seen as good fit, and values between .05 and .08 as an adequate fit, while values greater than .10 are unacceptable.

SRMR. The SRMR (Jöreskog & Sörbom, 1981) is a badness-of-fit measure based on the fitted residuals. Any standardised residual whose absolute value is greater than 1.96 or 2.58 can help detect the source of bad fit (Schermelleh-Engel et al., 2003). According to Hu and Bentler (1995), SRMR values less than .05 can be interpreted as a good fit, and values less than .10 as an acceptable fit.

3.6.5 Missing Data and Parameter Estimation Technique

Longitudinal surveys like the MCS are susceptible to missing data, as the multiple informants and data collection sweeps can cause many possible types of non-response/attrition (Laird, 1988; Plewis, 2007b). Inappropriate techniques for handling missing data can bias the parameter estimates, standard errors, and model fit (Arbuckle, 1996; Davey, 2005; Little & Rubin, 1989). When data are missing completely at random (MCAR)

or at random (MAR), the likelihood of biased estimates is negligible, but not when data are missing not at random (NMAR) (Little & Rubin, 1989; Newman, 2003; Rubin, 1976).

There are nine missingness patterns in the current dataset: 549 cohort members (89.7% of the working sample) had complete data for all variables, with nine (1.47%) missing some teacher-report data at sweep 4; 43 (7.03%) missing BAS Word Reading and/or teacher-report data at sweep 4; and 11 (1.80%) missing maternal education data. Little's MCAR test (Little, 1988) suggested MCAR for sweep-4 teacher-report missing data, $\chi^2_{[7]} = 10.8$, $p = .15$. Missing BAS data occurred because some parents chose the Welsh language assessment for their children instead of the BAS subtest in English, whereas missing maternal education data were subject to data recoding (i.e., overseas qualifications = missing). Therefore, all missingness was assumed to be MAR, as they appeared to be independent of the missing data themselves (Newman, 2003).

The MAR pattern allowed for the implementation of maximum likelihood (ML) as a parameter estimation technique, which seeks the variances-covariances that make the observed covariances most likely (Allison, 2003; Myung, 2003). Compared to other estimation techniques (e.g., listwise/pairwise deletion, single/multiple imputation), ML appears to be the optimal method for handling missingness for SEM (Allison, 2003; Duncan, Duncan, & Li, 1998), due to its sufficiency, consistency, efficiency, and parameterisation invariance (Myung, 2003). Further, in this study, a robust ML (MLR) estimation was required and adopted as the 'TYPE = COMPLEX' and 'WEIGHT' options were utilised in Mplus (Muthén & Muthén, 2017). The MLR computes robust chi-squares and standard errors, using a mean-adjusted likelihood ratio test statistic (Asparouhov & Muthén, 2005).

3.6.7 Data Distribution

ML assumes multivariate normality of variables (Allison, 2003; Ory & Mokhtarian, 2010), and can generate biased standard errors and incorrect test statistics given excessive multivariate kurtosis (Hu & Bentler, 1995). However, MLR is robust to violation of normality (Asparouhov & Muthén, 2005; Maydeu-Olivares, 2017), data distribution was therefore not a concern for this study.

Chapter 4: Results

This chapter presents the statistical results. First, some preliminary analyses will be reported. Second, results from the measurement model assessing the factorial validity of the latent constructs will be outlined. Third, results from the structural model will be presented, including the direct pathways between variables. Finally, indirect pathways from the mediation analyses will be interpreted.

Table 1

Descriptive Statistics of Family Background at Sweep 1 and HLE at Sweep 2

	N	*M (SE)*	Min	Max	Skewness	Kurtosis
Sweep 1						
Maternal education[a]	601	2.50 (0.06)	0	5	-0.19	-1.02
Family income[b]	612	0.77 (0.02)	0	1	-1.26	-0.42
Mother's literacy/numeracy skills						
Read storybooks[c]	612	0.94 (0.01)	0	1	-3.67	11.5
Fill out forms[c]	612	0.96 (0.01)	0	1	-4.84	21.4
Check change[c]	612	0.97 (0.01)	0	1	-6.07	34.8
Sweep 2						
Home tutoring						
Alphabets[d]	612	3.05 (0.11)	0	7	0.50	-1.10
Numbers/counting[d]	612	5.44 (0.10)	0	7	-0.96	-0.59
Songs/poems/rhymes[d]	612	5.71 (0.09)	0	7	-1.30	0.33
Home literacy practices						
Being read to[e]	612	5.56 (0.03)	2	6	-1.94	4.01
Library visit[f]	612	0.84 (0.05)	0	4	1.13	0.10
Watching TV/videos[g]	612	2.94 (0.03)	1	4	-0.24	-0.13
Painting/drawing[d]	612	5.25 (0.10)	0	7	-0.79	-0.84

Note. HLE = home literacy environment. *N* = number of observations. *M* = mean. *SE* = standard error. [a]0 (no formal qualifications) to 5 (higher/first degree). [b]0 (below poverty line) to 1 (above poverty line). [c]0 (low) to 1 (high). [d]0 (not at all) to 7 (7 times a week). [e]1 (not at all) to 6 (every day). [f]0(not at all) to 4 (once a week). [g]1 (not at all) to 4 (more than 3 hours).

Table 2

Descriptive Statistics of English Literacy and Language Skills at Ages 5, 7

	N	*M (SE)*	Min	Max	Skewness	Kurtosis
Age 5						
Vocabulary[a]	612	55.3 (0.44)	20	80	-0.07	0.45
Phonological awareness[b]	612	6.99 (0.10)	0	9	-1.05	-0.01
Print concept[b]	612	7.12 (0.08)	0	9	-0.92	0.43
Age 7						
Decoding[a]	569	109 (0.86)	55	145	-0.14	-0.19
Speaking/listening[c]	611	3.40 (0.04)	1	5	-0.11	-0.09
Reading[c]	597	3.31 (0.05)	1	5	-0.15	-0.53
Writing[c]	597	3.07 (0.04)	1	5	-0.04	-0.24

Note. [a]Standardised scores. [b]Raw summed scores of the number of attributes attained. [c]1 (well below average) to 5 (well above average). All variables were adopted from teacher report, apart from vocabulary and decoding being based on cognitive assessment results.

4.1 Preliminary Data Analysis

Tables 1 and 2 show the univariate descriptive statistics for the measures of family background, home literacy environment (HLE), and children's literacy and language skills, calculated from weighted data. Accordingly, the skewness and kurtosis values of all variables were in the acceptable range (Kline, 2015), not considering the dichotomous variables (e.g., maternal abilities), as skewness and kurtosis are normality measurements for continuous variables (Cain et al., 2017). All variables were standardised before further analyses. The correlation matrix for all variables, including the covariates is displayed in Appendix B.

The actual frequencies of most observed variables are shown in tables in Appendix C. In our sample, 432 families had income above 60% of national median around childbirth. Most mothers (33.8%) held a highest educational qualification at NVQ level 2 (i.e., GCSE grades A*-C; National Council for Vocational Qualifications, 1995), while the majority reported to have no difficulty performing the three daily tasks, ranging from 93.5% to 97.5% of the working sample (N = 612). In terms of home tutoring, most parents reported to teach numbers/counting (58.8%) and songs/poems/rhymes (65.4%) seven times a week, while

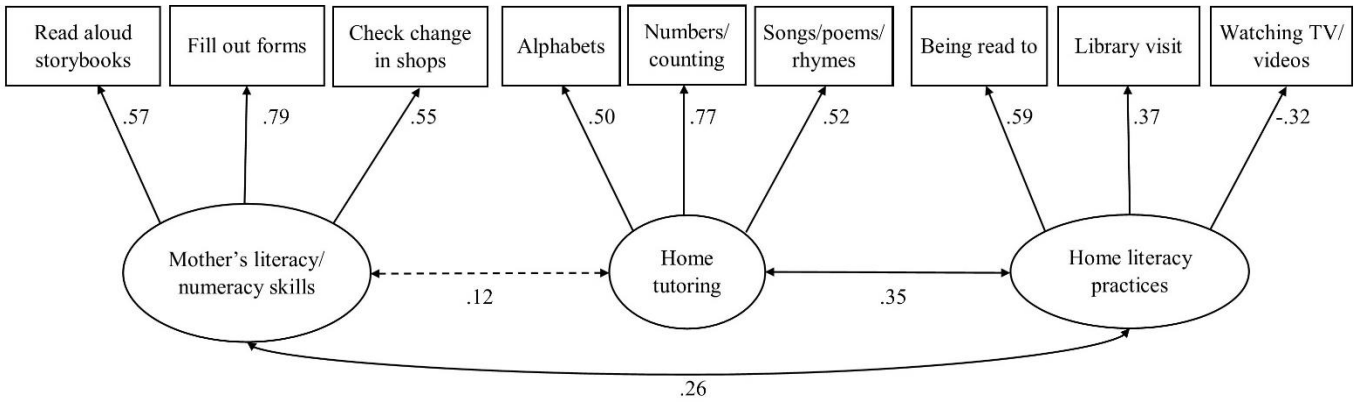
alphabets/letters were taught less frequently, as 22.4% of the parents reported to do so seven times a week, and 21.9% reported once or twice a week. Home literacy practices were also performed frequently, as 66.7% of the cohort members were read to every day, and 19.9% were read to several times a week. Meanwhile, 53.4% of the cohort members engaged in painting/drawing at home seven times a week. Regarding television and videos watching, most cohort members (56.9%) spent up to an hour per day on this. However, library visit seemed rare as most children were never taken to the library (60.1%). At school entry, 40.0% of the cohort members were reported by teachers to have attained all nine attributes on the 'Linking Sounds and Letters' subscale (23.3% attaining 7-8 attributes), while 27.3% of them had attained all print concept attributes (35.3% attaining 7-8 attributes). Two years later, large proportions of cohort members were reported by teachers to have average speaking/listening (44.8% of 611 children), reading (35.5% of 597 children), and writing skills (41.8% of 597 children).

4.2 Measurement Model

The measurement model for maternal self-reported literacy/numeracy skills, home tutoring, and home literacy practices was estimated using confirmatory factor analysis (CFA). One parental interview item (the frequency of children's painting/drawing) loaded weakly onto the 'home literacy practices' latent variable and attenuated the model fit, MLR $\chi^2_{[32]}$ = 88.5, $p < .001$; CFI = .87; RMSEA = .054, 90% CIs = [.04, .07]; SRMR = .045. Model modification by removing this item resulted in better model fit. Therefore, only three interview items (frequency of children being read to at home, frequency of library visit, frequency of children's watching television/video) were retained as indicators of home literacy practices. The final CFA model (Figure 4) showed good fit to the data and supported the independence of home tutoring and home literacy practices as two components of HLE. Meanwhile, the two HLE constructs were positively correlated ($p < .001$), so were maternal skills and home literacy practices ($p < .01$). There was a weak positive trend between maternal skills and home tutoring, which was however not significant ($p < .07$). Notably, the indicator 'watching television/videos' loaded negatively onto the home literacy practices latent construct, meaning that increases in home literacy practices were associated with reduced time spent on watching television/videos.

Figure 4

Confirmatory Factor Analysis of Mother's Literacy/numeracy Skill Variables around Childbirth and Home Literacy Environment Variables at Age 3



Note. TV = television. The latent variables are in ovals and the observed variables are in rectangles. Dashed line represents non-significant relationship. Values on single-headed arrows are standardised factor loadings. Values on double-headed arrows are standardised factor correlations. MLR $\chi^2_{[24]} = 42.1$, $p = .01$; CFI = .95; RMSEA = .035, 90% CIs = [.02, .05]; SRMR = .036.

4.3 Structural Model

Table 3 shows the covariance matrix for all the latent and observed variables used to construct the longitudinal structural model. The final model is presented in Figure 5 and shows excellent model fit. Overall, the model explained 38.5% of the variance in children's decoding ability as assessed by BAS Word Reading, 32.7% of the variance in teacher-reported speaking/listening, 39.0% of the variance in teacher-reported reading, and 41.9% of the variance in teacher-reported writing.

Table 3

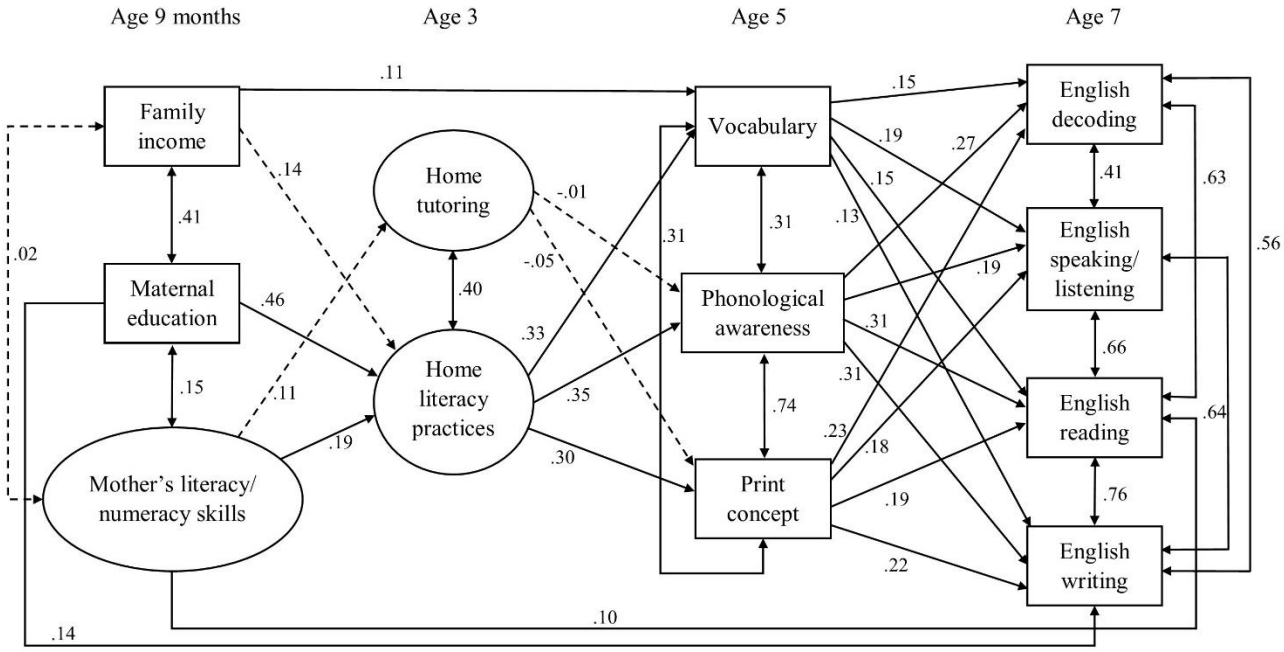
Covariance Matrix for Latent and Observed Variables

Variables	1	2	3	4	5	6	7	8	9	10	11	12
Latent variables												
1. Maternal skills	0.02**											
2. Home literacy practices	0.02**	0.16***										
3. Home tutoring	0.02	0.18**	1.63***									
Observed variables												
4. Maternal education	0.03**	0.32***	0.04	2.04***								
5. Family income	0.00	0.06***	0.00	0.29***	0.18***							
Early skills at age 5												
6. BAS-NV	0.07	1.38***	1.35*	3.00***	0.98***	100***						
7. PA	0.07**	0.40***	0.40*	0.98***	0.23***	9.59***	5.80***					
8. PC	0.04**	0.26***	0.14	0.77***	0.17***	7.10***	3.46***	3.34***				
Later outcomes at age 7												
9. BAS-WR	0.43**	2.56***	2.69	8.08***	1.63***	70.2***	25.1***	18.4***	354***			
10. Speaking/listening	0.02**	0.10**	0.15*	0.33***	0.09***	3.35***	1.04***	0.78***	10.1***	0.81***		
11. Reading	0.03**	0.12***	0.17*	0.42***	0.11***	3.82***	1.44***	1.04***	15.2***	0.74***	1.13***	
12. Writing	0.02***	0.10**	0.13	0.41***	0.10***	3.43***	1.38***	1.00***	13.4***	0.68***	0.90***	0.99***

Note. BAS-NV = BAS Naming Vocabulary; PA = phonological awareness; PC = print concept; BAS-WR = BAS Word Reading. *$p < .05$. **$p < .01$. ***$p < .001$.

Figure 5

Longitudinal Structural Equation Model Predicting English Decoding, Speaking/listening, Reading, and Writing at Age 7



Note. Vocabulary and English decoding are based on child performance data, the other variables at ages 5 and 7 are based on teacher report. The latent variables are in ovals and the observed variables are in rectangles. Dashed lines represent non-significant pathways. Values on single-headed arrows are standardised regression weights. Values on double-headed arrows are standardised correlation coefficients. MLR $\chi^2_{[99]} = 167$, $p < .001$; CFI = .98; RMSEA = .034, 90% CIs = [.02, .04]; SRMR = .032.

4.3.1 Direct Pathways from Family Background

Family income did not predict home literacy practices (*SE* = .08, *p* = .08), inconsistent with the hypothesis. Instead, it directly predicted children's vocabulary at age 5 (*SE* = 06, *p* = .047), indicating that in our sample, the relationships between family background and the precursor skills were not fully mediated by HLE. Maternal education predicted the home literacy practice latent construct as hypothesised (*SE* = .07, *p* < .001). Notably, it also directly predicted teacher-reported English writing at age 7 (*SE* = .06, *p* = .03). Mother's self-reported literacy/numeracy skills predicted home literacy practices (*SE* = .7, *p* = .004) but not home tutoring (*SE* = .06, *p* = .08), yet it directly predicted a teacher-reported children's skill assessed 7 years later, the English reading skill (*SE* = .04, *p* = .004). The direct pathways

from family background to later literacy and language outcomes suggested that full mediations via HLE and precursor skills were not present in our sample.

4.3.2 Direct Pathways from HLE

The associations between home tutoring and the precursor skills at age 5 are not as hypothesised, as pathways from home tutoring to both teacher-reported phonological awareness (*SE* = .06, *p* = .92) and print concept (*SE* = 06, *p* = .42) were non-significant. Conversely, the associations between home literacy practices and the early skills are consistent with our hypothesis, as it significantly predicted children's vocabulary (*SE* = .09, *p* < .001), phonological awareness (*SE* = .13, *p* = .006), and print concept (*SE* = .11, *p* =.007).

4.3.3 Direct Pathways from Early Literacy and Language Skills

Each direct pathway from the early literacy and language skills to the later outcomes was significant, with *p*-values ranging from <.001 to .002. Likewise, the correlations among the skill variables were significant when tested concurrently, at both ages 5 and 7 (*p*s < .001). These findings confirmed to our hypotheses regarding the concurrent and longitudinal relationships among the literacy and language skills at sweeps 3 and 4.

4.4 Mediation Analysis

The indirect effects of family background and HLE variables on children's literacy and language skills were also estimated, using bias-corrected bootstrapped confidence intervals (CIs) (Preacher & Hayes, 2004), with 1,000 bootstrap samples. Specifically, we measured the indirect pathways from family background around childbirth to precursor skills at age 5, and to the later literacy and language outcomes at age 7. The indirect pathways from HLE at age 3 to later outcomes via the mediation of precursor skills at age 5 were also examined. Standardised coefficients, standard errors, *p*-values, and bias-corrected bootstrapped 95% CIs for the indirect effects are presented in Tables 4-6.

Indirect effects with bootstrapped CIs that do not include zero are considered as significant (Preacher & Hayes, 2004). As shown in Tables 4-6, there are discrepancies in determining the significance of the indirect effects between *p*-values and CIs. Namely, many pathways had bootstrapped CIs that did not include zero, indicating significant effects, whilst the *p*-values suggested non-significant effects. This was probably due to the asymmetrical CIs, which is common in statistical practice, and associated with a skewed sampling

distribution (MacKinnon & Dwyer, 1993; Stone & Sobel, 1990). In this case, the CIs should be used to infer significance, as they account for the asymmetry in the distribution (Mallinckrodt et al., 2006).

4.4.1 Indirect Pathways from Family Background

Table 4

Indirect Effects of Family Background on Early Literacy and Language Skills at Age 5

Path	*β*	*SE*	*p*	BC bootstrapped 95% CI
ME → LP → NV	.15	.06	.009	[.05, .28]
FI → LP → NV	.05	.04	.199	[-.01, .13]
MS → LP → NV	.06	.04	.086	[.01, .15]
ME → LP → PA	.16	.09	.068	[.05, .38]
FI → LP → PA	.05	.04	.247	[-.01, .16]
MS → LP → PA	.07	.05	.157	[.01, .18]
MS → HT → PA	.00	.01	.943	[-.03, .02]
ME → LP → PC	.14	.08	.070	[.04, .33]
FI → LP → PC	.04	.04	.249	[-.01, .14]
MS → LP → PC	.06	.04	.156	[.01, .16]
MS → HT → PC	-.01	.01	.619	[-.03, .01]

Note. SE = standard error; BC = bias-corrected; CI = confidence interval; ME = maternal education; LP = home literacy practices; NV = BAS Naming Vocabulary; PA = teacher-reported phonological awareness; PC = teacher-reported print concept; FI = family income; MS = maternal self-reported literacy/numeracy skills; HT = home tutoring.

Table 5

Indirect Effects of Family Background on Later Literacy and Language Outcomes at Age 7

Path	*β*	*SE*	*p*	BC bootstrapped 95% CI
ME → LP → NV → WR	.02	.01	.053	[.01, .05]
ME → LP → PA → WR	.04	.03	.164	[.01, .12]
ME → LP → PC → WR	.03	.02	.164	[.01, .08]
FI → LP → NV → WR	.01	.01	.264	[.00, .02]
FI → LP → PA → WR	.01	.01	.290	[.00, .05]
FI → LP → PC → WR	.01	.01	.307	[.00, .03]
MS → LP → NV → WR	.01	.01	.147	[.00, .03]
MS → LP → PA → WR	.02	.02	.230	[.00, .06]
MS → HT → PA → WR	.00	.00	.948	[-.01, .00]
MS → LP → PC → WR	.01	.01	.221	[.00, .04]
MS → HT → PC → WR	.00	.00	.683	[-.01, .00]
ME → LP → NV → SL	.03	.01	.038	[.01, .06]

ME → LP → PA → SL	.03	.02	.205	[.01, .09]
ME → LP → PC → SL	.03	.02	.165	[.01, .07]
FI → LP → NV → SL	.01	.01	.239	[.00, .03]
FI → LP → PA → SL	.01	.01	.358	[.00, .04]
FI → LP → PC → SL	.01	.01	.323	[.00, .03]
MS → LP → NV → SL	.01	.01	.147	[.00, .03]
MS → LP → PA → SL	.01	.01	.272	[.00, .04]
MS → HT → PA → SL	.00	.00	.948	[-.01, .00]
MS → LP → PC → SL	.01	.01	.250	[.00, .04]
MS → HT → PC → SL	.00	.00	.668	[-.01, .00]
ME → LP → NV → RE	.02	.01	.065	[.01, .05]
ME → LP → PA → RE	.05	.04	.172	[.01, .13]
ME → LP → PC → RE	.03	.02	.249	[.01, .07]
FI → LP → NV → RE	.01	.01	.291	[.00, .02]
FI → LP → PA → RE	.02	.02	.322	[.00, .06]
FI → LP → PC → RE	.01	.01	.334	[.00, .03]
MS → LP → NV → RE	.01	.01	.181	[.00, .03]
MS → LP → PA → RE	.02	.02	.251	[.00, .07]
MS → HT → PA → RE	.00	.00	.946	[-.01, .01]
MS → LP → PC → RE	.01	.01	.292	[.00, .03]
MS → HT → PC → RE	.00	.00	.737	[-.01, .00]
ME → LP → NV → WT	.02	.01	.075	[.00, .05]
ME → LP → PA → WT	.05	.04	.169	[.01, .13]
ME → LP → PC → WT	.03	.03	.257	[.01, .08]
FI → LP → NV → WT	.01	.01	.291	[.00, .02]
FI → LP → PA → WT	.02	.02	.314	[.00, .06]
FI → LP → PC → WT	.01	.01	.356	[.00, .03]
MS → LP → NV → WT	.01	.01	.197	[.00, .03]
MS → LP → PA → WT	.02	.02	.240	[.00, .07]
MS → HT → PA → WT	.00	.00	.948	[-.01, .01]
MS → LP → PC → WT	.01	.01	.300	[.00, .04]
MS → HT → PC → WT	.00	.00	.744	[-.01, .00]

Note. WR = BAS Word Reading; SL = teacher-reported speaking/listening; RE = teacher-reported reading; WT = teacher-reported writing.

According to Table 4, maternal education and self-reported literacy/numeracy skills had indirect effects on all the early literacy and language skills at age 5 (vocabulary, phonological awareness, print concept), via home literacy practices but not home tutoring. Moreover, all indirect pathways from family income to the precursor skills were non-significant but it directly predicted children's vocabulary (see Section 4.3.1). Similarly, Table 5 shows that home literacy practices and all precursor skills at age 5 were significant

mediators for the relationships between family background and later outcomes at age 7, whereas indirect pathways involving home tutoring were all non-significant.

4.4.2 Indirect Pathways from HLE

Table 6 suggests that, the effects of home literacy practices on later literacy and language outcomes at age 7 (decoding, teacher-reported speaking/listening, reading, writing) were fully mediated by vocabulary, teacher-reported phonological awareness, and print concept at age 5. Conversely, home tutoring was not indirectly predictive of any outcome variable, as all the CIs included zero.

Table 6

Indirect Effects of HLE on Later Literacy and Language Outcomes at Age 7

Path	*β*	*SE*	*p*	BC bootstrapped 95% CI
HT → PA → WR	.00	.02	.938	[-.06, .03]
HT → PC → WR	-.01	.02	.554	[-.05, .02]
LP → NV → WR	.05	.02	.032	[.01, .10]
LP → PA → WR	.10	.06	.092	[.03, .23]
LP → PC → WR	.07	.04	.093	[.02, .16]
HT → PA → SL	.00	.02	.939	[-.04, .02]
HT → PC → SL	-.01	.02	.555	[-.05, .01]
LP → NV → SL	.06	.03	.019	[.02, .13]
LP → PA → SL	.07	.04	.140	[.02, .17]
LP → PC → SL	.06	.04	.108	[.01, .14]
HT → PA → RE	.00	.03	.937	[-.07, .04]
HT → PC → RE	-.01	.02	.613	[-.05, .01]
LP → NV → RE	.05	.02	.041	[.01, .10]
LP → PA → RE	.11	.07	.100	[.04, .26]
LP → PC → RE	.06	.04	.160	[.01, .14]
HT → PA → WT	.00	.03	.938	[-.07, .04]
HT → PC → WT	-.01	.02	.618	[-.05, .02]
LP → NV → WT	.04	.02	.052	[.01, .10]
LP → PA → WT	.11	.07	.095	[.03, .26]
LP → PC → WT	.07	.05	.168	[.01, .16]

Chapter 5: Discussion

In this study, we aimed to investigate the roles of family background and home literacy environment (HLE) in predicting the development of early English literacy and language skills among 612 children in Wales, using data from a UK national longitudinal survey, the Millennium Cohort Study (MCS). So far, the MCS has followed the cohort members from birth in Year 2000 to sweep 7 at age 17. This study extracted data from sweeps 1 to 4, especially measures of family background around childbirth, the HLE at age 3, early literacy and language skills at age 5 (vocabulary, phonological awareness, print concept), and later literacy and language outcomes at age 7 (decoding, speaking/listening, reading, writing). The HLE was conceptualised as home tutoring (i.e., direct instructions on areas that may foster literacy and language development) and home literacy practices (i.e., informal parent-child interactions that reflect parents' resources, values, and attitudes towards children's literacy and language learning) (after Nag et al., 2019). Based on previous findings and particularly Hamilton et al.'s (2016) model, we examined the direct and mediated relationships between family background, HLE and the literacy and language measures using a longitudinal structural equation model, when variance shared with child sex, season of birth, and household single-parent status was controlled.

This chapter will interpret the statistical results from Chapter 4 and analyse how the findings fit in with our research questions and existing literature. It will then conclude with the strengths, limitations, and implications of this dissertation.

5.1 Family Background Predicting Early Literacy and Language Development

5.1.1 Associations between Family Background and HLE

Our model indicates that home literacy practices were significantly predicted by maternal education and self-reported literacy/numeracy skills, but not family income, partially supporting our hypothesis. These findings show that in the Wales sample, regardless of family income, mothers who were more-educated and those who can perform daily literacy- and numeracy-related activities (i.e., read children's storybooks, fill out forms, calculate change) easily were more likely to carry out casual parent-child interactions relevant for children's literacy and language development, such as book reading and library visit. The differential effects of maternal education and family income confirm to the notion that the relationship between these two variables tends to be strong but not monotonic, as

mothers with some education, despite living in low-income families, may also have positive attitudes towards children's literacy learning and attempt to provide available resources (Hartas, 2011, Nag et al., 2019). The positive association between maternal education and home literacy practices, however, has been supported by a wide range of literature in different countries (e.g., US: Carneiro et al., 2013; India: Kalia & Reese, 2009; Netherlands: Van Steensel, 2006).

The links between maternal literacy/numeracy skills and the two aspects of HLE were exploratory, as little research has been done using mother's self-reported skills. The findings suggest that maternal skills predicted home literacy practices but not home tutoring, which is in fact consistent with Puglisi et al.'s (2017) research findings using mother's language assessment scores. This implies that mother's self-reported skills in daily activities, although not formally assessed, may reliably represent their literacy and numeracy skills.

5.1.2 Family Background Predicting Early Literacy and Language Development

The mediation analyses present some important findings on the relationships between family background and early literacy and language skills at age 5 (vocabulary, phonological awareness, print concept). All the indirect pathways from family income to the precursor skills were non-significant, instead, it directly predicted children's vocabulary assessed by BAS Naming Vocabulary. Contrarily, the indirect effects of maternal education and literacy/numeracy skills on all the precursor skills were observed, but only via home literacy practices, not home tutoring. Thus, the hypothesis that the relationships between family background and early skills would be fully mediated by the HLE was partially supported. The direct relationship between family income and children's vocabulary may emphasise that living in poverty since birth could have a persistently negative effect on children's oral language by school entry, which is consistent with the notion that early SES-related differences can create gaps in literacy and language outcomes (Fernald et al., 2013; Hart & Risley, 1995; Walker et al., 1994). However, the mediation of home literacy practices on the pathways from maternal education and skills to children's precursor skills denotes that having a mother with a high educational level, and/or high abilities to perform daily tasks is beneficial for a child's literacy and language development. This is implemented as the mother provides a more resourceful HLE, through casual activities that can increase children's exposure to print, in line with Hamilton et al.'s (2016) finding and some other research (e.g., Forget-Dubois et al., 2009; Foster et al., 2005). The different performances of pathways from

family income and maternal education again highlight that the relationship between these two variables is not monotonic (Hartas, 2011), supporting the decision to treat them as separate variables instead of indicators for a construct.

Regarding the predictions of later literacy and language outcomes, some direct relationships should be noted. Contradicting the hypothesis on full mediations, a significant direct pathway from maternal education around childbirth to teacher-reported children's writing ability at age 7 was observed, so was the pathway from mother's self-reported literacy/ numeracy skills to teacher-reported children's reading ability at age 7. These findings may reflect a genetic influence on children's literacy and language development. As suggested by van Bergen et al. (2017), parents with high literacy skills tend to have high levels of educational attainment and highly literate children. Although the parents in van Bergen et al.'s (2017) study were directly assessed on reading fluency, while the abilities of parents in our study were self-reported, the two studies revealed similar trends in the relationships between maternal and children's literacy skills. Additionally, all indirect pathways from the family background variables to later literacy and language outcomes via home literacy practices at age 3 and precursor skills at age 5 were significant. The findings of several direct and indirect pathways suggest partial mediations on the relationships between family background and children's outcomes, instead of full mediations, consistent with Hamilton et al.'s (2016) findings.

Altogether, this shows that family background is highly likely to have a persistent impact on children's literacy and language development, both directly and indirectly, even when the measurements were obtained far apart (i.e., family background around childbirth, later literacy outcomes at age 7). Particularly, even mother's self-reported literacy/numeracy skills and teacher-reported children' skills were significantly and directly linked, consistent with findings from studies that formally tested mother's language skills (Puglisi et al., 2017). This implies that not only mother's assessment performance, but also their abilities to carry out daily literacy- and numeracy-related activities can explain some variance in children's literacy and language development. The direct pathways from mother-related variables to later child outcomes provide evidence for the notion that a substantial proportion of children's literacy and language skills is inherited (Christopher et al., 2015; de Zeeuw et al., 2015).

5.2 HLE Predicting Early Literacy and Language Development

5.2.1 HLE Predicting Early Skills at Age 5

HLE is considered more proximal to children's literacy and language development than family background (Hamilton et al., 2016; Inoue et al., 2018). However, opposite to what was hypothesised, home tutoring was found to be non-predictive of teacher-reported phonological awareness and print concept. This finding is also inconsistent with the home literacy model which posits that home tutoring can predict code-related literacy skills, including print concept (Martini & Sénéchal, 2012; Sénéchal & LeFevre, 2002), as well as previous findings of home tutoring significantly predicting phonological awareness (e.g., Hamilton et al., 2016; Inoue et al., 2018; Liu et al., 2018). Contrarily, home literacy practices significantly predicted all three early literacy and language skills, consistent with our hypothesis and some preceding literature (e.g., Korat et al., 2013; Van Steensel, 2006).

In fact, these findings are in line with Hartas's (2011) regression analysis on the whole MCS population, where only the frequency of book reading, but not home tutoring, was reported to be associated with teacher-reported literacy and language. Hartas (2011) proposed that this was because shared book reading played a more vital role in fostering children's literacy and language development than any form of home tutoring, and that home tutoring alone may not reduce the SES disparity, even though parents from diverse SES backgrounds engaged in home tutoring equally frequently. This argument may be relevant for our study on the Wales sample of the MCS, although Hartas (2011) analysed the whole MCS population and did not differentiate between the subscales of the 'Communication, Language and Literacy' section in the teacher questionnaire.

We recognise that home literacy practices, especially shared book reading as a prototypical aspect of it, possess strong associations with children's early literacy and language skills. Nonetheless, the assertion by the home literacy model that direct home tutoring predicts the development of code-related skills has also been supported by many researchers (e.g., Hood et al., 2008; Inoue et al., 2018; Stephenson et al., 2008). To explain the non-significant relationships between home tutoring and phonological awareness and print concept in our sample, the attention may be drawn to the operationalisation of the home tutoring construct. Specifically, by including the frequencies of parental instruction on alphabets, numbers/counting, and songs/poems/rhymes, the quantity and diversity of home tutoring were captured, but not the quality of these teaching activities. For instance, the

context of teaching, teaching materials and the difficulty level of these materials were all unknown. Leseman and De Jong (1998) have emphasised the importance of the quality of parental teaching in children's literacy and language development, as they believed simple exposure was insufficient to facilitate child development. This notion was reflected in our analysis, as most parents reported to engage in home tutoring frequently (e.g., 7 times a week), which had not translated into children's abilities. Hence, other characteristics of home tutoring may play a role, such as its quality (Sénéchal & LeFevre, 2002). However, the quality measures were not collected in the MCS (Hansen, 2014), so this explanation may be valid but cannot be systematically tested in our sample.

5.2.2 HLE Predicting Later Outcomes at Age 7

HLE also indirectly predicted children's literacy and language outcomes at age 7. Multiple indirect pathways from home literacy practices to later outcomes (decoding assessed by BAS Word Reading, teacher-reported speaking/listening, reading, and writing) were observed, via full mediations of the precursor skills at age 5. This is not surprising as similar relationships have been revealed in existing literature (e.g., Hamilton et al., 2016; Inoue et al., 2018). However, the home tutoring construct performed differently and all the indirect pathways from this construct to the later outcomes were non-significant, inconsistent with our prediction and some previous findings (e.g., Hamilton et al., 2016; Inoue et al., 2018). Accordingly, home literacy practices at age 3 directly predicted the precursor skills at age 5, and indirectly predicted the later outcomes at age 7, whereas home tutoring did not predict children's literacy and language development.

Collectively, our results suggest that in the Wales sample, the home literacy practices construct is a strong predictor of children's literacy and language development, directly predicting the precursor skills at age 5 and indirectly predicting the later outcomes at age 7. Therefore, engagement in activities such as shared book reading and library visit can benefit children's vocabulary, phonological awareness, and print concept skills, which in turn can benefit later skills such as reading and writing. Conversely, home tutoring did not play an important role, which will be further discussed later.

5.3 Precursor Skills Predicting Later Literacy and Language Outcomes

5.3.1 Concurrent Correlations among Literacy and Language Skills

At age 5, children's vocabulary as an oral language skill, phonological awareness as a meta-linguistic skill, and print concept as a conceptual literacy skill significantly correlated with each other. This is consistent with Sénéchal et al.'s (2001) conceptualisation of early literacy, as well as many previous findings across countries and languages (e.g., Burgess & Lonigan, 1998; Chung et al., 2021; Dickinson et al., 2003; Lonigan et al., 2008; McBride-Chang et al., 2005; Uchikoshi, 2019). Significant correlations among the later literacy and language skills at age 7 (word decoding assessed by BAS Word Reading, teacher-reported speaking/listening, reading, and writing) were also revealed, in line with the notion that speaking, listening, reading, and writing are closely and concurrently related in primary school years (Berninger & Abbott, 2010). The hypothesis regarding significant concurrent relationships is thus supported.

Notably, there are differences in the strength of correlations between and within the two sweeps. Specifically, at age 5, the correlation between the two teacher-report variables (phonological awareness, print concept) was stronger (r = .74) than when the cognitive assessment variable (vocabulary) was correlated with a teacher-report variable (phonological awareness: r = .31; print concept: r = .31). These results may be explained Duncan et al.'s (2007) finding from a comparative analysis on six longitudinal datasets that the correlations between academic skills were the lowest when they were obtained from different informants. However, this difference in correlation strength seems to have faded away at age 7, as the six correlations among the outcome variables were equivalently strong, with coefficients ranging from .56 to .76, except for the correlation between decoding and teacher-reported speaking/listening (r = .41). Hence, Duncan et al.'s (2007) finding may not be sufficient to account for age-7 results. One may then argue with evidence from Meisels et al.'s (2001) study, that the weaker correlations at age 5 were subject to teachers' low levels of familiarity with new students, which had been boosted by age 7. However, whether the cohort members were rated by the same teachers at both sweeps was unknown. Alternatively, this may be explained by the developmental sequence of oral language (e.g., vocabulary) and code-related skills (e.g., phonological awareness, print concept). Indeed, oral language skills have been demonstrated to develop much earlier than code-related skills, so that children tend to bring oral language skills to formal schooling, but not necessarily the code-related skills (Storch &

Whitehurst, 2002; Whitehurst & Lonigan, 1998). Therefore, the weaker concurrent relationships between vocabulary and code-related skills at age 5 possibly reflects their developmental sequence. However, these skills were tested concurrently and obtained from different informants in our study, meaning that no definite claim can be made.

5.3.2 Developmental Relationships among Literacy and Language Skills

Our model indicates that all the associations between early and later literacy and language skills were significant, even after controlling for the distal family background and proximal HLE variables, as hypothesised. This means that in our Wales sample, children who scored high in precursor skills (vocabulary, teacher-reported phonological awareness, and print concept) tended to also score high in later literacy and language skills (decoding, teacher-reporting speaking/listening, reading, and writing). This finding agrees with the body of existing literature which states that early literacy and language skills are developmental precursors to later skills (e.g., Foorman et al., 2015; Kendeou et al., 2009; Kim et al., 2015; Muter et al., 2004; Storch & Whitehurst, 2002).

However, it is worth noting that most previous studies focused on reading comprehension as the only outcome variable and directly examined children's skills using cognitive assessments, whereas this study involved more outcome variables reported by teachers. Indeed, several studies have highlighted the validity of teacher report in representing children's emergent literacy skills and academic achievement, by correlating teacher ratings with children's performance outcomes (e.g., Beswick et al., 2005; Cabell et al., 2009; Meisels et al., 2001). In this study, subject to data availability, it was unfeasible to compare teacher ratings against children's actual performance. Therefore, although similar trends of early literacy and language skills predicting later outcomes were revealed in this study, consistent with previous findings, further research using direct assessments is warranted.

5.4 Strengths of this Dissertation

The strengths of the dissertation will be discussed in terms of its theoretical contributions to the body of existing literature, and its methodological advantages.

5.4.1 Theoretical Contributions

Findings from the current study add to the existing literature in several ways. First, the inclusion of family background variables (family income, maternal education, mother's self-reported literacy/numeracy skills) as single measurement constructs instead of indicators for a unitary construct is valuable, as each variable was associated with children's literacy and language skills in different ways. Meanwhile, apart from replicating previous findings, our study emphasises the persistent impact of maternal education and literacy/numeracy skills on child development. Significant associations were revealed even when they and the outcome variables were tested seven years apart, whereas previous research with similar findings involved variables that were tested one year apart only (e.g., Puglisi et al., 2017). Therefore, this may be considered as a striking finding, supporting the view on the heritability of children's literacy and language skills (Christopher et al., 2015; de Zeeuw et al., 2015).

Second, the conceptualisation of HLE as home tutoring and home literacy practices, as well as the inclusion of activities in other domains than reading and writing, may be useful in future research. The inclusion of more activities from more domains reflects the multidimensional nature of HLE. Activities related to art, music, and numeracy for example, have been suggested to foster literacy and language development (e.g., Heath & Wolf, 2005; Paquette & Rieg, 2008; Purpura & Napoli, 2015). In this study, the lower frequency of watching television/videos as an indicator for the home literacy practices construct contributed to the prediction of children's literacy and language development. Although the home tutoring latent construct was not a significant predictor, we argue this is probably not because teaching on non-literacy domains was included, but because more characteristics of teaching (e.g., quality, difficulty level) should be assessed other than frequency. A broader scope in conceptualising the HLE construct should be considered by future research.

Third, this study contributes to a better understanding of the English literacy and language development of children in Wales, a stable English-Welsh bilingual community. Under exposure to both English and Welsh from an early age, the developmental trajectories of English literacy and language skills in our Wales sample broadly confirmed to findings from English monolinguals (e.g., Hamilton et al., 2016). Considering this, our findings can be used to inform parental practices in Wales and probably other linguistic communities, which will be explicated later.

5.4.2 Methodological Advantages

One methodological advantage is related to the use of the MCS dataset. The MCS is a nationally representative longitudinal survey that contains information on over 4,000 variables (Hansen, 2014), among which the variables on family background around childbirth, HLE at age 3, and literacy and language skills at ages 5 and 7 allowed this study to be conducted. As the first birth cohort study to include all four countries of UK (England, Wales, Scotland, Northern Ireland) (Hansen, 2014, p. 5), the MCS allowed us to investigate a relatively under-researched population – children in Wales. Moreover, its longitudinal nature is advantageous as it suited the research interests of this study to explore both concurrent and developmental relationships.

Additionally, this study employed advanced statistical techniques to analyse the Wales sample and test the research hypotheses. The oversampling, non-response, and attrition were accounted for by incorporating weights and strata in model specification, ensuring the representativeness of our results at least within the Wales context (Korn & Graubard, 1995). Meanwhile, missing data was handled using robust maximum likelihood, a parameter estimator that is optimal for handling missingness in structural equation modelling (SEM), compared to other estimation techniques such as listwise deletion and multiple imputation (Allison, 2003; Myung, 2003). Also, the factorial validity of all latent constructs was supported by a confirmatory factor analysis. Further, the direct and mediated relationships among variables were tested using SEM, a statistical technique believed to outperform some traditional techniques like regression (MacCallum & Austin, 2000). SEM allowed multiple relationships to be computed simultaneously (Hoyle, 1995; Quinn & Wagner, 2018), which was advantageous especially for the current study, as a combination of latent and observed variables were included, and multiple mediated relationships needed to be estimated. Altogether, we believe the statistical techniques have enhanced the validity, reliability, and representativeness of our findings.

5.5 Limitations and Suggestions for Future Research

Limitations of this study include three aspects, the questionable validity of some variables, lack of consideration of home and school languages, and limitations of the statistical methods.

5.5.1 Questionable Validity of Some Variables

The validity of some variables may be challenged. First, one may question the validity of teacher-report variables in representing children's literacy and language skills. Although the validity of teacher judgements has been supported by various research, which reported significant correlations between teacher ratings and child performance (e.g., Begeny & Buchanan, 2010; Beswick et al., 2005; Cabell et al., 2009; Meisels et al., 2001), we recognise that these correlations cannot be tested in our study due to data unavailability. Also, at sweep 3, the teachers in Wales were not systematically trained to fill out the questionnaires equivalent to the Foundation Stage Profile in England (Johnson, 2008), so the validity of their answers was not guaranteed. Further, at sweep 4, teachers were asked a singular question concerning children's abilities in speaking/listening, reading, and writing respectively, which may be problematic as the constructs underlying teacher ratings were not clear. Specifically, how teachers interpreted each question was unknown. For example, reading might have been interpreted as either decoding or comprehension, or both (Hoover & Gough, 1990), which would have been treated as distinct constructs in cognitive assessments (e.g., Inoue et al., 2018). This potentially limits the generalisability of our findings and the possibility of cross-study comparisons, as teachers were not trained to answer the sweep-3 questionnaire, and the underlying constructs of the sweep-4 measures were not clear. Thus, conclusions cannot be drawn on the developmental relationships associated with these teacher-reported skills. Correspondingly, future research may only claim the validity of teacher report after examining its consistency with child performance and may design more rigorous questions to capture teacher judgements.

Moreover, the construct of home tutoring needs further inspection. This is not to say the validity of its indicators is questionable, instead, we argue that some other characteristics of home tutoring may be necessary when assessing the relationships between home tutoring and children's literacy and language development. Indeed, most parents in our sample engaged in the three types of home tutoring (alphabets, numbers/counting, and songs/poems/rhymes) frequently, which did not result in better performance in every child. Therefore, it seems reasonable to argue that simple exposure to home tutoring may be insufficient to close the gaps derived from early disparities, and some other characteristics of home tutoring may be crucial, such as the quality, duration, and context of teaching (Leseman & De Jong, 1998; Sénéchal & LeFevre, 2002). This could not be explored in this study, as we

failed to find relevant variables that tapped these characteristics of home tutoring activities in the MCS, but future research may achieve this by collecting primary data.

5.5.2 Lack of Considerations of Home and School Languages

One may argue that knowing the language spoken at home is essential for research on bilingual children, as the timing of acquisition of a bilingual's two languages is associated with the amount of home language inputs in each language (Eilers et al., 2002). Namely, even though the Wales sample was analysed as a whole, home language may have impacted on children's English literacy and language development and should probably have been included in the model as a covariate. However, due to data security reasons, access to the exact language spoken in each household was restricted, and the only variable on home language informed how much English was spoken at home in relation to all other languages. This was insufficient for our research, as knowledge on the proportion of Welsh spoken at home is also necessary for research on children in Wales (Strand & Lindorff, 2021). Therefore, this home language variable in the MCS was decided not to be included.

Similarly, school language has been considered vital, as children's literacy and language development is potentially facilitated when home and school languages are the same (Lesaux et al., 2006; Nag et al., 2019). The school language was reported by teachers at both sweeps 3 and 4, but susceptible to high levels of missingness. Hence, neither home nor school language was included in analysis. Nevertheless, future research may consider comparing the effects of family background and HLE on the literacy and language development between Wales children whose home and school languages are consistent and whose two languages are inconsistent. This may help reveal the effects of home and school languages on the relationships between distal family background, proximal HLE, and literacy and language development, among children living in this stable English-Welsh bilingual community.

5.5.3 Limitations of Statistical Methods

A major limitation of the analysis concerns the direction and nature of the relations between constructs and their indicators. Specifically, maternal skills, and the two HLE constructs were treated as latent constructs, yet it may be more sensible to treat them as formative constructs (e.g., Stadler et al., 2021). According to the concept of latent construct, constructs produce behaviour that is captured by the indicators, such that changes in the latent

construct should result in changes in every indicator (Bollen & Lennox, 1991; McGrew, 2009). Accordingly, increases in the construct of maternal literacy/numeracy skills for instance, should cause increases in all three types of abilities (read children's books, fill out forms, calculate change). However, it may be more sensible to conceptualise that increase in each ability can cause changes in maternal skills as a construct, which confirms to the definition of formative construct (Blalock, 1971). Indeed, we had attempted to form the three constructs as formative. However, due to limited resources and time, it was unlikely to grasp how to include multiple formative constructs in a complex structural equation model, as virtually all previous examples presented only one formative construct and usually as an independent variable (e.g., Dolean et al., 2019; Roberts & Thatcher, 2009). This difficulty may be overcome by future research.

SEM also has limitations. For example, the evaluation of a model relies on the goodness-of-fit indices, yet excellent model fit does not mean the model depicts the reality. McDonald and Ho (2002) argued that other model specifications may produce equally good model fit, and that model misspecifications can be masked by seemingly good fitting indices. Nonetheless, by reporting the measurement model and structural model separately as recommended by McDonald and Ho (2002), our model may be less susceptible to this limitation of SEM. Nonetheless, if there were no time constraints, we would have compared the current model with alternative models to verify whether the current model exhibited the highest model fit.

5.6 Implications

The findings of this study suggest that frequent engagement in home literacy practices is beneficial for children's literacy and language development in a long-term. Apart from shared reading which has been widely recognised as a strong predictor, in our study, other activities (e.g., library visit) have added to the effects of home literacy practices. Therefore, where possible, parents are recommended to provide a diverse range of experiences to facilitate the growth of children's literacy and language skills. Based on the current study, however, specific recommendations about whether parents should attempt to teach such skills at home would be premature. In this study, only the frequency of home tutoring was known, but not the duration, quality, or difficulty level of teaching. As most parents reported to conduct home tutoring frequently, differences in competencies were not detected between

children whose parents involved in home tutoring and those whose parents did not. Hence, suggestions about the extent and depth of home tutoring cannot be made.

Nevertheless, training programmes should be designed to offer instructions to parents, regarding how to best support child development via HLE. Our results stress the persistent impact of family background on children's literacy and language development, especially the mother-related variables (i.e., maternal education, literacy/numeracy skills). Consequently, mothers from not so affluent backgrounds may need external assistance in providing a resourceful HLE, to compensate for low levels of education and/or difficulties in performing daily literacy/numeracy tasks. This may help to narrow the gaps in literacy and language skills derived from early disparities. We believe these implications can be generalised to other contexts outside Wales, as the effects of family background and HLE have been reported across countries and languages.

There are some implications for future research. Specifically, we suggest that family background should not be treated as a unitary construct, as each attribute of it tends to perform differently in relationships with children's literacy and language skills. Furthermore, we suggest future research to always measure more aspects of HLE, other than activities related to reading and writing.

5.7 Conclusion

As a secondary data analysis on the MCS, this study examined the developmental relationships between family background around childbirth, HLE at age 3, and the English literacy and language development between ages 5 and 7, among 612 children in Wales, after controlling for child sex, season of birth, and household single-parent status. We replicated and extended the theoretical model adopted from Hamilton et al.'s (2016) study. Results from a longitudinal structural equation model indicated that the effects of family background on later children's literacy and language outcomes at age 7 were partially mediated by home literacy practices (not home tutoring), and the early skills at age 5. Similarly, the effects of home literacy practices, but not home tutoring, on later child outcomes at age 7 were fully mediated by the precursor skills at age 5. Direct pathways were also revealed between family income and children's English vocabulary, between maternal education and English writing, and between maternal skills and English reading.

This study added to the literature about predictors of English literacy and language development among Welsh children. We identified a stable impact of family background around childbirth (family income, maternal education, and self-reported literacy/numeracy skills) on children's literacy and language skills during early primary school, emphasising the possible genetic influences. This study also contributed to the understanding of HLE, by suggesting that home-based activities in other domains should also be assessed to better predict children's literacy and language development, other than activities related to reading and writing. Future research may thus adopt a broader scope in operationalising HLE and treat each attribute of family background as separate variables. However, future research is also needed to investigate how other characteristics of home tutoring (e.g., quality), and the inconsistency in home and school languages affect children's literacy and language development. When teacher-report measures are used, rigorously designed questionnaire items are crucial, and their validity should be assessed via correlations with child performance.

References

Aarnoutse, C., Van Leeuwe, J., Voeten, M., & Oud, H. (2001). Development of decoding, reading comprehension, vocabulary and spelling during the elementary school years. *Reading and Writing, 14*(1), 61-89. https://doi.org/10.1023/A:1008128417862

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 545-557. https://doi.org/10.1037/0021-843X.112.4.545

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14*(5), 255-259. https://doi.org/10.1111/j.0963-7214.2005.00376.x

Aram, D., Korat, O., Saiegh-Haddad, E., Arafat, S. H., Khoury, R., & Elhija, J. A. (2013). Early literacy among Arabic-speaking kindergartners: The role of socioeconomic status, home literacy environment and maternal mediation of writing. *Cognitive Development*, *28*(3), 193-208. https://doi.org/10.1016/j.cogdev.2012.10.003

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 243-277). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Arya, D. J., McClung, N. A., Maul, A., & Cunningham, A. E. (2014). The effects of early home literacy environments on fourth-grade literacy achievement: an international comparison. *International Journal of Quantitative Research in Education*, *2*(1), 1-16. https://doi.org/10.1504/IJQRE.2014.060971

Asparouhov, T., & Muthén, B. (2005). Multivariate statistical modeling with survey data. In *Proceedings of the Federal Committee on Statistical Methodology (FCSM) research conference* (pp. 14-16). http://www.statmodel.com/download/2005FCSM_Asparouhov_Muthen_IIA.pdf

Backlund, P. (1985). Essential speaking and listening skills for elementary school students. *Communication Education, 34*(3), 185-195. https://doi.org/10.1080/03634528509378606

Begeny, J. C., & Buchanan, H. (2010). Teachers' judgments of students' early literacy skills measured by the Early Literacy Skills Assessment: Comparisons of teachers with and without assessment administration experience. *Psychology in the Schools, 47*(8), 859-868. https://doi.org/10.1002/pits.20509

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. https://doi.org/10.1037/0033-2909.107.2.238

Berninger, V. W., & Abbott, R. D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in grades 1, 3, 5, and 7. *Journal of Educational Psychology, 102*(3), 635-651. https://doi.org/10.1037/a0019319

Berninger, V. W., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and reading: Connections between language by hand and language by eye. *Journal of Learning Disabilities, 35*(1), 39-56. https://doi.org/10.1177/002221940203500104

Beswick, J. F., Willms, J. D., & Sloat, E. A. (2005). A comparative study of teacher ratings of emergent literacy skills and student performance on a standardized measure. *Education, 126*(1), 116-138. https://www.researchgate.net/profile/Elizabeth-Sloat-2/publication/50876676_A_Comparative_Study_of_Teacher_Ratings_of_Emergent_Literacy_Skills_and_Student_Performance_on_a_Standardized_Measure/links/558a953208aeae8413bcc804/A-Comparative-Study-of-Teacher-Ratings-of-Emergent-Literacy-Skills-and-Student-Performance-on-a-Standardized-Measure.pdf

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage?. *Applied Psycholinguistics, 24*(1), 27-44. https://doi.org/10.1017/S014271640300002X

Blalock, H. M. (1971). Causal Models Involving Unobserved Variables in Stimulus-Response Situations. In: H. M. Blalock (Ed.), *Causal Models in the Social Sciences* (pp. 335-347). Chicago, IL: Aldine.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*(2), 305-314. https://doi.org/10.1037/0033-2909.110.2.305

Bornstein, M. H., Hahn, C.-S., Suwalsky, J. T. D., & Haynes, O. M. (2003). Socioeconomic status, parenting and child development: The Hollingshead four-factor index of social status and the socioeconomic index of occupations. In M. H. Bornstein & R. H. Bradley (Eds.), *Socioeconomic status, parenting, and child development: Monographs in parenting* (pp. 29-82). Mahwah, NJ: Lawrence Erlbaum.

Boslaugh, S. (2007). *Secondary data sources for public health: A practical guide*. New York, NY: Cambridge University Press.

Bowey, J. A. (1994). Phonological sensitivity in novice readers and nonreaders. *Journal of Experimental Child Psychology, 58*(1), 134-159. https://doi.org/10.1006/jecp.1994.1029

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: SAGE Publications.

Bryne, B. (2009). *Structural equation modeling using AMOS: Basic concepts, applications, and programming* (2nd ed.). London: Routledge.

Burchinal, M. R., Peisner-Feinberg, E., Pianta, R., & Howes, C. (2002). Development of academic skills from preschool through second grade: Family and classroom predictors of developmental trajectories. *Journal of School Psychology, 40*(5), 415-436. https://doi.org/10.1016/S0022-4405(02)00107-3

Burgess, S. (2005). The preschool home literacy environment provided by teenage mothers. *Early Child Development and Care, 175*(3), 249-258. https://doi.org/10.1080/0300443042000266303

Burgess, S. R., & Lonigan, C. J. (1998). Bidirectional relations of phonological sensitivity and prereading abilities: Evidence from a preschool sample. *Journal of Experimental Child Psychology, 70*(2), 117-141. https://doi.org/10.1006/jecp.1998.2450

Burgess, S. R., Hecht, S. A., & Lonigan, C. J. (2002). Relations of the home literacy environment (HLE) to the development of reading-related abilities: A one-year longitudinal study. *Reading Research Quarterly, 37*(4), 408-426. https://doi.org/10.1598/RRQ.37.4.4

Cabell, S. Q., Justice, L. M., Zucker, T. A., & Kilday, C. R. (2009). Validity of teacher report for assessing the emergent literacy skills of at-risk preschoolers. *Language, Speech, and Hearing Services in Schools, 40*(2), 161-173. https://doi.org/10.1044/0161-1461(2009/07-0099)

Cain, M. K., Zhang, Z., & Yuan, K. H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods, 49*(5), 1716-1735. https://doi.org/10.3758/s13428-016-0814-1

Carneiro, P., Meghir, C., & Parey, M. (2013). Maternal education, home environments, and the development of children and adolescents. *Journal of the European Economic Association, 11*(suppl_1), 123-160. https://doi.org/10.1111/j.1542-4774.2012.01096.x

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences, 110*(28), 11278-11283. https://doi.org/10.1073/pnas.1309518110

Chaney, C. (1992). Language development, metalinguistic skills, and print awareness in 3-year-old children. *Applied Psycholinguistics, 13*(4), 485-514. https://doi.org/10.1017/S0142716400005774

Chaney, C. (1994). Language development, metalinguistic awareness, and emergent literacy skills of 3-year-old children in relation to social class. *Applied Psycholinguistics, 15*(3), 371-394. https://doi.org/10.1017/S0142716400004501

Chaney, C. (1998). Preschool language and metalinguistic skills are links to reading success. *Applied Psycholinguistics, 19*(3), 433-446. https://doi.org/10.1017/S0142716400010250

Chazan-Cohen, R., Raikes, H., Brooks-Gunn, J., Ayoub, C., Pan, B. A., Kisker, E. E., Roggman, L., & Fuligni, A. S. (2009). Low-income children's school readiness:

Parent contributions over the first five years. *Early Education and Development, 20*(6), 958-977. https://doi.org/10.1080/10409280903362402

Christian, K., Morrison, F. J., & Bryant, F. B. (1998). Predicting kindergarten academic skills: Interactions among child care, maternal education, and family literacy environments. *Early Childhood Research Quarterly, 13*(3), 501-521. https://doi.org/10.1016/S0885-2006(99)80054-4

Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B., DeFries, J. C., Wadsworth, S. J., Willcutt, E., & Olson, R. K. (2015). Genetic and environmental etiologies of the longitudinal relations between prereading skills and reading. *Child Development, 86*(2), 342-361. https://doi.org/10.1111/cdev.12295

Chung, S. C., Geva, E., Chen, X., & Deacon, S. H. (2021). Do We 'Laugh' or 'La8gh'? Early Print Knowledge and Its Relation to Learning to Read in English and French. *Scientific Studies of Reading*, 1-15. https://doi.org/10.1080/10888438.2020.1863970

Connelly, R. (2013). *Interpreting Test Scores. Millennium Cohort Study Data Note 2013/01*. London: Centre for Longitudinal Studies.

Connelly, R., & Platt, L. (2014). Cohort profile: UK millennium Cohort study (MCS). *International Journal of Epidemiology, 43*(6), 1719-1725. https://doi.org/10.1093/ije/dyu001

Crawford, C., Dearden, L., & Greaves, E. (2013). *When you are born matters: evidence for England* (IFS Report No. R80). London: Nuffield Foundation.

Dale, A., Arbor, S., & Procter, M. (1988). *Doing secondary analysis*. London: Unwin Hyman.

Dale, A., Wathan, J., & Wiggins, V. (2008). Secondary analysis of quantitative data sources. In P. Alasuutari, L. Bickman, & J. Brannen (Eds.) *The Sage handbook of social research methods* (pp. 520-535). London: SAGE Publications.

Dauber, S. L., & Epstein, J. L. (1993). Parents' attitudes and practices of involvement in inner-city elementary and middle schools. In N. F. Chavkin (Ed.), *Families and*

schools in a pluralistic society (pp. 53-71). Albany, NY: State University of New York Press.

Davey, A. (2005). Issues in evaluating model fit with missing data. *Structural Equation Modeling, 12*(4), 578-597. https://doi.org/10.1207/s15328007sem1204_4

De Jong, P. F., & Leseman, P. P. (2001). Lasting effects of home literacy on reading achievement in school. *Journal of School Psychology, 39*(5), 389-414. https://doi.org/10.1016/S0022-4405(01)00080-2

de Zeeuw, E. L., de Geus, E. J., & Boomsma, D. I. (2015). Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement. *Trends in Neuroscience and Education, 4*(3), 69-76. https://doi.org/10.1016/j.tine.2015.06.001

Dickinson, D. K., & Snow, C. E. (1987). Interrelationships among prereading and oral language skills in kindergartners from two social classes. *Early Childhood Research Quarterly, 2*(1), 1-25. https://doi.org/10.1016/0885-2006(87)90010-X

Dickinson, D. K., McCabe, A., Anastasopoulos, L., Peisner-Feinberg, E. S., & Poe, M. D. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology, 95*(3), 465-481. https://doi.org/10.1037/0022-0663.95.3.465

Dolean, D., Melby-Lervåg, M., Tincas, I., Damsa, C., & Lervåg, A. (2019). Achievement gap: socioeconomic status affects reading development beyond language and cognition in children facing poverty. *Learning and Instruction, 63*, 101218. https://doi.org/10.1016/j.learninstruc.2019.101218

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, 56*(8), 848-856. https://doi.org/10.1111/jcpp.12378

Dulay, K. M., Cheung, S. K., Reyes, P., & McBride, C. (2019). Effects of parent coaching on Filipino children's numeracy, language, and literacy skills. *Journal of Educational Psychology, 111*(4), 641-662. https://doi.org/10.1037/edu0000315

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428-1446. https://doi.org/10.1037 /0012-1649.43.6.1428

Duncan, T. E., Duncan, S. C., & Li, F. (1998). A comparison of model-and multiple imputation-based approaches to longitudinal analyses with partial missingness. *Structural Equation Modeling: A Multidisciplinary Journal, 5*(1), 1-21. https://doi.org/10.1080/10705519809540086

Dunsmuir, S., & Blatchford, P. (2004). Predictors of writing competence in 4-to 7-year-old children. *British Journal of Educational Psychology, 74*(3), 461-483. https://doi.org/10.1348/0007099041552323

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*(3), 250-287. https://doi.org/10.1598/RRQ.36.3.2

Eilers, R. E., Oller, D. K., & Cobo-Lewis, A. B. (2002). Bilingualism and cultural assimilation in Miami Hispanic children. In: D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 43-63). Clevedon, England: Multilingual Matters.

Elliott, C. D., Smith, P., & McCulloch, K. (1996). *British Ability Scales Second Edition (BAS II). Administration and Scoring Manual*. London: Nelson.

Ermisch, J., & Francesconi, M. (2001). Family matters: Impacts of family background on educational attainments. *Economica, 68*(270), 137-156. https://doi.org/10.1111/1468-0335.00239

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. *Social Science Research, 33*(3), 464-497. https://doi.org/10.1016/j.ssresearch.2003.08.001

Farver, J. A. M., Xu, Y., Eppe, S., & Lonigan, C. J. (2006). Home environments and young Latino children's school readiness. *Early Childhood Research Quarterly, 21*(2), 196-212. https://doi.org/10.1016/j.ecresq.2006.04.008

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

Fitzsimons, E., Haselden, L., Smith, K., Gilbert, E., Calderwood, L., AgaliotiSgompou, V., Veeravalli, S., Silverwood, R., Ploubidis, G. (2020). *Millennium Cohort Study Age 17 Sweep (MCS7): User Guide*. London: Centre for Longitudinal Studies, University College of London.

Florit, E., Roch, M., Altoè, G., & Levorato, M. C. (2009). Listening comprehension in preschoolers: The role of memory. *British Journal of Developmental Psychology, 27*(4), 935-951. https://doi.org/10.1348/026151008X397189

Foorman, B. R., Herrera, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). The structure of oral language and reading and their relation to comprehension in Kindergarten through Grade 2. *Reading and Writing, 28*(5), 655-681. https://doi.org/10.1007/s11145-015-9544-5

Forget-Dubois, N., Dionne, G., Lemelin, J. P., Pérusse, D., Tremblay, R. E., & Boivin, M. (2009). Early child language mediates the relation between home environment and school readiness. *Child Development, 80*(3), 736-749. https://doi.org/10.1111/j.1467-8624.2009.01294.x

Foster, M. A., Lambert, R., Abbott-Shim, M., McCarty, F., & Franze, S. (2005). A model of home learning environment and social risk factors in relation to children's emergent literacy and social outcomes. *Early Childhood Research Quarterly, 20*(1), 13-36. https://doi.org/10.1016/j.ecresq.2005.01.006

Frijters, J. C., Barron, R. W., & Brunello, M. (2000). Direct and mediated influences of home literacy and literacy interest on prereaders' oral vocabulary and early written language skill. *Journal of Educational Psychology, 92*(3), 466-477. https://doi.org/10.1037/0022-0663.92.3.466

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1-25. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x

Gan, L., & Chong, S. (1998). The rhythm of language: fostering oral and listening skills in Singapore pre-school children through an integrated music and language arts program. *Early Child Development and Care, 144*(1), 39-45. https://doi.org/10.1080/0300443981440105

Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition, 12*(2), 213-237. https://doi.org/10.1017/S1366728909004015

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. *American Psychologist, 50*(1), 24-37. https://doi.org/10.1037/0003-066X.50.1.24

Gillon, G. T. (2004). *Phonological Awareness: From Research to Practice*. New York, NY: Guilford Press.

Goldin-Meadow, S., Levine, S. C., Hedges, L. V., Huttenlocher, J., Raudenbush, S. W., & Small, S. L. (2014). New evidence about language and cognitive development based on a longitudinal study: Hypotheses for intervention. *American Psychologist, 69*(6), 588-599. https://doi.org/10.1037/a0036886

Goswami, U. (2001). Early phonological development and the acquisition of literacy. In S.B. Neuman & D.K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 111-125). New York, NY: The Guilford Press.

Graham, S., & Harris, K. R. (2005). Improving the writing performance of young struggling writers: Theoretical and programmatic research from the center on accelerating student learning. *The Journal of Special Education, 39*(1), 19-33. https://doi.org/10.1177/00224669050390010301

Gray, J. C., & Simmonds, N. (2009). *Millennium Cohort Study Sweep 3 Teacher Survey in Northern Ireland, Scotland and Wales Technical Report*. London: Centre for Longitudinal Studies, Institute of Education.

Gromko, J. E. (2005). The effect of music instruction on phonemic awareness in beginning readers. *Journal of Research in Music Education, 53*(3), 199-209. Doi: https://doi.org/10.1177/002242940505300302

Hamilton, L. G., Hayiou-Thomas, M. E., Hulme, C., & Snowling, M. J. (2016). The home literacy environment as a predictor of the early literacy development of children at family-risk of dyslexia. *Scientific Studies of Reading, 20*(5), 401-419. https://doi.org/10.1080/10888438.2016.1213266

Hansen, K. (2014). *Millennium Cohort Study: A guide to the datasets* (8th Ed.). London: Centre for Longitudinal Studies, Institute of Education.

Hart, B., & Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H. Brookes.

Hartas, D. (2011). Families' social backgrounds matter: Socio-economic factors, home learning and young children's language, literacy and social outcomes. *British Educational Research Journal, 37*(6), 893-914. https://doi.org/10.1080/01411926.2010.506945

Heath, S. B., & Wolf, S. (2005). Focus in creative learning: Drawing on art for language development. *Literacy, 39*(1), 38-45. https://doi.org/10.1111/j.1741-4350.2005.00396.x

Hecht, S. A., Burgess, S. R., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2000). Explaining social class differences in growth of reading skills from beginning kindergarten through fourth-grade: The role of phonological awareness, rate of access, and print knowledge. *Reading and Writing, 12*(1), 99-128. https://doi.org/10.1023/A:1008033824385

Hemmerechts, K., Agirdag, O., & Kavadias, D. (2017). The relationship between parental literacy involvement, socio-economic status and reading literacy. *Educational Review, 69*(1), 85-101. https://doi.org/10.1080/00131911.2016.1164667

Hill-Clarke, K., & Robinson, N. R. (2004). It's as easy as A-B-C and do-re-mi: Music, rhythm, and rhyme enhance children's literacy skills. *Young Children*, *59*(5), 91-95. https://www.proquest.com/scholarly-journals/as-easy-b-c-do-re-mi-music-rhythm-rhyme-enhance

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, *49*(1), 4-14. https://doi.org/10.1037/a0027238

Holm, A., & Dodd, B. (1996). The effect of first written language on the acquisition of English literacy. *Cognition, 59*(2), 119-147. https://doi.org/10.1016/0010-0277(95)00691-5

Hood, M., Conlon, E., & Andrews, G. (2008). Preschool home literacy practices and children's literacy development: A longitudinal analysis. *Journal of Educational Psychology, 100*(2), 252-271. https://doi.org/10.1037/0022-0663.100.2.252

Hooper, S. R., Roberts, J. E., Nelson, L., Zeisel, S., & Kasambira Fannin, D. (2010). Preschool predictors of narrative writing skills in elementary school children. *School Psychology Quarterly*, *25*(1), 1-12. https://doi.org/10.1037/a0018329

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127-160. https://doi.org/10.1007/BF00401799

Hox, J. J., & Bechger, T. M. (1998). An Introduction to Structural Equation Modeling. *Family Science Review*, *11*, 354-373. https://dspace.library.uu.nl/handle/1874/23738

Hoyle, R. H. (1995). The structural equation modeling approach: Basic concepts and fundamental issues. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 1-15). London: SAGE Publications.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). London: SAGE Publications.

Huston, A., McLoyd, V. C., & Garcia-Coll, C. (1994). Children and poverty: Issues in contemporary research. *Child Development, 65*, 275-282. https://doi.org/10.1111/j.1467-8624.1994.tb00750.x

Hyman, H. H. (1972). *Secondary Analysis of Sample Surveys: Principles, Procedures and Potentialities*. London: John Wiley & Sons.

Inoue, T., Georgiou, G. K., Parrila, R., & Kirby, J. R. (2018). Examining an extended home literacy model: The mediating roles of emergent literacy skills and reading fluency. *Scientific Studies of Reading*, *22*(4), 273-288. https://doi.org/10.1080/10888438.2018.1435663

Janssen, C., Segers, E., Mcqueen, J. M., & Verhoeven, L. (2017). Transfer from implicit to explicit phonological abilities in first and second language learners. *Bilingualism: Language and Cognition, 20*(4), 795-812. https://doi.org/10.1017/S1366728916000523

Jeong, J., McCoy, D. C., & Fink, G. (2017). Pathways between paternal and maternal education, caregivers' support for learning, and early child development in 44 low- and middle-income countries. *Early Childhood Research Quarterly, 41*, 136-148. https://doi.org/10.1016/j.ecresq.2017.07.001

Johnson, J. (2008). *Millennium third survey follow-up: A guide to the school assessment datasets*. London: Centre for Longitudinal Studies, Institute of Education.

Johnston, M. P. (2017). Secondary data analysis: A method of which the time has come. *Qualitative and Quantitative Methods in Libraries, 3*(3), 619-626. http://www.qqml-journal.net/index.php/qqml/article/view/169

Jordan, G. E., Snow, C. E., & Porche, M. V. (2000). Project EASE: The effect of a family literacy project on kindergarten students' early literacy skills. *Reading Research Quarterly, 35*(4), 524-546. https://doi.org/10.1598/RRQ.35.4.5

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relationships by maximum likelihood and least squares methods* (Research Report 81−8). Uppsala, Sweden: University of Uppsala, Department of Statistics.

Justice, L. M., & Ezell, H. K. (2001). Word and print awareness in 4-year-old children. *Child Language Teaching and Therapy, 17*(3), 207-225. https://doi.org/10.1177/026565900101700303

Kalia, V., & Reese, E. (2009). Relations between Indian children's home literacy environment and their English oral language and literacy skills. *Scientific Studies of Reading, 13*(2), 122-145. https://doi.org/10.1080/10888430902769517

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101*(4), 765-778. https://doi.org/10.1037/a0015956

Kielcolt, K. J., & Nathan, L. E. (1986). *Secondary Analysis of Survey Data*. London: SAGE Publications.

Kim, Y. S., Al Otaiba, S., & Wanzek, J. (2015). Kindergarten predictors of third grade writing. *Learning and Individual Differences, 37*, 27-37. https://doi.org/10.1016/j.lindif.2014.11.009

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Press.

Korat, O., Arafat, S. H., Aram, D., & Klein, P. (2013). Book reading mediation, SES, home literacy environment, and children's literacy: Evidence from Arabic-speaking families. *First Language, 33*(2), 132-154. https://doi.org/10.1177/0142723712455283

Korn, E. L., & Graubard, B. I. (1995). Examples of differing weighted and unweighted estimates from a sample survey. *The American Statistician, 49*(3), 291-295. https://doi.org/10.1080/00031305.1995.10476167

Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence?. *Child Development, 70*(6), 1373-1400. https://doi.org/10.1111/1467-8624.00101

Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine, 7*(1-2), 305-315. https://doi.org/10.1002/sim.4780070131

Lamb, S. J., & Gregory, A. H. (1993). The relationship between music and reading in beginning readers. *Educational Psychology, 13*(1), 19-27. https://doi.org/10.1080/0144341930130103

Lee, J. A. C., & Al Otaiba, S. (2015). Socioeconomic and gender group differences in early literacy skills: A multiple-group confirmatory factor analysis approach. *Educational Research and Evaluation, 21*(1), 40-59. https://doi.org/10.1080/13803611.2015.1010545

Lefebvre, P., Trudeau, N., & Sutton, A. (2011). Enhancing vocabulary, print awareness and phonological awareness through shared storybook reading with low-income preschoolers. *Journal of Early Childhood Literacy, 11*(4), 453-479. https://doi.org/10.1177/1468798411416581

Lesaux, N. K., Koda, K., Siegel, L. S. & Shanahan, T. (2006). Development of literacy. In: D. August, & T. Shanahan (Eds.), *Developing literacy in second-language learners* (pp. 75-122). Mahway, NJ: Erlbaum.

Leseman, P. P., & De Jong, P. F. (1998). Home literacy: Opportunity, instruction, cooperation and social-emotional quality predicting early reading achievement. *Reading Research Quarterly, 33*(3), 294-318. https://doi.org/10.1598/RRQ.33.3.3

Liebeskind, K. G., Piotrowski, J. T., Lapierre, M. A., & Linebarger, D. L. (2014). The home literacy environment: Exploring how media and parent–child interactions are associated with children's language production. *Journal of Early Childhood Literacy, 14*(4), 482-509. https://doi.org/10.1177/1468798413512850

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198-1202. https://doi.org/10.1080/01621459.1988.10478722

Little, R. J., & Rubin, D. B. (1989). The analysis of social science data with missing values. *Sociological Methods & Research*, *18*(2-3), 292-326. https://doi.org/10.1177/0049124189018002004

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford Press.

Liu, C., Georgiou, G. K., & Manolitsis, G. (2018). Modeling the relationships of parents' expectations, family's SES, and home literacy environment with emergent literacy skills and word reading in Chinese. *Early Childhood Research Quarterly, 43*, 1-10. https://doi.org/10.1016/j.ecresq.2017.11.001

Lonigan, C. J., Anthony, J. L., Phillips, B. M., Purpura, D. J., Wilson, S. B., & McQueen, J. D. (2009). The nature of preschool phonological processing abilities and their relations to vocabulary, general cognitive abilities, and print knowledge. *Journal of Educational Psychology*, *101*(2), 345-358. https://doi.org/10.1037/a0013837

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), 596-613. https://doi.org/10.1037/0012-1649.36.5.596

Lonigan, C. J., Burgess, S. R., Anthony, J. L., & Barker, T. A. (1998). Development of phonological sensitivity in 2- to 5-year-old children. *Journal of Educational Psychology*, *90*(2), 294-311. https://doi.org/10.1037/0022-0663.90.2.294

Lonigan, C. J., Schatschneider, C., Westberg, L., & National Early Literacy Panel. (2008). Identification of children's skills and abilities linked to later outcomes in reading, writing, and spelling. In *National Early Literacy Panel, Developing early literacy, report of the National Early Literacy Panel: A scientific synthesis of early literacy development and implications for intervention* (pp. 55-106). Washington, DC: National Center for Family Literacy.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*(1), 201-226. https://doi.org/10.1146/annurev.psych.51.1.201

Machida, S., Taylor, A. R., & Kim, J. (2002). The role of maternal beliefs in predicting home learning activities in Head Start families. *Family Relations, 51*(2), 176-184. https://doi.org/10.1111/j.1741-3729.2002.00176.x

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17*(2), 144-158. https://doi.org/10.1177/0193841X9301700202

Mallinckrodt, B., Abraham, W. T., Wei, M., & Russell, D. W. (2006). Advances in testing the statistical significance of mediation effects. *Journal of Counseling Psychology, 53*(3), 372-378. https://doi.org/10.1037/0022-0167.53.3.372

Martini, F., & Sénéchal, M. (2012). Learning literacy skills at home: Parent teaching, expectations, and child interest. *Canadian Journal of Behavioural, 44*(3), 210-221. https://doi.org/10.1037/a0026758

Mason, J. M., & Stewart, J. P. (1990). Emergent literacy assessment for instructional use in kindergarten. In L. M. Morrow & J. K. Smith (Eds.), *Assessment for Instruction in Early Literacy* (pp. 155-175). Englewood Cliffs, NJ: Prentice-Hall.

Maydeu-Olivares, A. (2017). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(3), 383-394. https://doi.org/10.1080/10705511.2016.1269606

McBride-Chang, C., Cho, J. R., Liu, H., Wagner, R. K., Shu, H., Zhou, A., Cheuk, C. S-M., & Muse, A. (2005). Changing models across cultures: Associations of phonological awareness and morphological structure awareness with vocabulary and word recognition in second graders from Beijing, Hong Kong, Korea, and the United States. *Journal of Experimental Child Psychology, 92*(2), 140-160. https://doi.org/10.1016/j.jecp.2005.03.009

McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*(1), 64-82. https://doi.org/10.1037/1082-989X.7.1.64

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin, 107*(2), 247-255. https://doi.org/10.1037/0033-2909.107.2.247

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1-10. https://doi.org/10.1016/j.intell.2008.08.004

Meisels, S. J., Bickel, D. D., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgments: A validity study of a curriculum-embedded performance assessment in kindergarten to grade 3. *American Educational Research Journal, 38*(1), 73-95. https://doi.org/10.3102/00028312038001073

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*(1), 3-19. https://doi.org/10.1037/0022-0663.91.1.3

Molfese, V. J., Beswick, J., Molnar, A., & Jacobi-Vessels, J. (2006). Alphabetic skills in preschool: A preliminary study of letter naming and letter writing. *Developmental Neuropsychology, 29*(1), 5-19. https://doi.org/10.1207/s15326942dn2901_2

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*(5), 665-681. https://doi.org/10.1037/0012-1649.40.5.665

Muthén, L.K., & Muthén, B.O. (2017). *Mplus User's Guide* (8th Ed.). Los Angeles, CA: Muthén & Muthén.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47*(1), 90-100. https://doi.org/10.1016/S0022-2496(02)00028-7

Nag, S., Vagh, S. B., Dulay, K. M., & Snowling, M. J. (2019). Home language, school language and children's literacy attainments: A systematic review of evidence from low-and middle-income countries. *Review of Education, 7*(1), 91-150. https://doi.org/10.1002/rev3.3130

National Council for Vocational Qualifications (NCVQ). (1995). *NVQ Criteria and Guidance*. London: NCVQ.

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods, 6*(3), 328-362. https://doi.org/10.1177/1094428103254673

O'Herron, P., & Siebenaler, D. (2007). The intersection between vocal music and language arts instruction: A review of the literature. *Update: Applications of Research in Music Education, 25*(2), 16-26. https://doi.org/10.1177/87551233070250020103

Ory, D. T., & Mokhtarian, P. L. (2010). The impact of non-normality, sample size and estimation technique on goodness-of-fit measures in structural equation modeling: evidence from ten empirical models of travel behavior. *Quality & Quantity, 44*(3), 427-445. https://doi.org/10.1007/s11135-008-9215-6

Pace, A., Luo, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics, 3*, 285-308. https://doi.org/10.1146/annurev-linguistics-011516-034226

Panizzon, M. S., Vuoksimaa, E., Spoon, K. M., Jacobson, K. C., Lyons, M. J., Franz, C. E., Xian, H., Vasilopoulos, T., & Kremen, W. S. (2014). Genetic and environmental influences on general cognitive ability: Is ga valid latent construct?. *Intelligence, 43*, 65-76. https://doi.org/10.1016/j.intell.2014.01.008

Paquette, K. R., & Rieg, S. A. (2008). Using music to support the literacy development of young English language learners. *Early Childhood Education Journal, 36*(3), 227-232. https://doi.org/10.1007/s10643-008-0277-9

Park, H. (2008). Home literacy environments and children's reading performance: A comparative study of 25 countries. *Educational Research and Evaluation, 14*(6), 489-505. https://doi.org/10.1080/13803610802576734

Perfetti, C. A. (1985). *Reading skills*. New York, NY: Oxford University Press.

Piacentini, M. (2014). *Measuring income inequality and poverty at the regional level in OECD countries* (OECD Working Paper No. 58). Paris: OECD Publishing. http://dx.doi.org/10.1787/5jxzf5khtg9t-en

Plewis, I. (2007a). *The Millennium Cohort study: Technical report on sampling* (4th Ed.). London: Centre for Longitudinal Studies, Institute of Education.

Plewis, I. (2007b). Non-response in a birth cohort study: the case of the Millennium Cohort Study. *International Journal of Social Research Methodology, 10*(5), 325-334. https://doi.org/10.1080/13645570701676955

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*(4), 717-731. https://doi.org/10.3758/BF03206553

Puglisi, M. L., Hulme, C., Hamilton, L. G., & Snowling, M. J. (2017). The home literacy environment is a correlate, but perhaps not a cause, of variations in children's language and literacy development. *Scientific Studies of Reading, 21*(6), 498-514. https://doi.org/10.1080/10888438.2017.1346660

Pullen, P. C., & Justice, L. M. (2003). Enhancing phonological awareness, print awareness, and oral language skills in preschool children. *Intervention in School and Clinic, 39*(2), 87-98. https://doi.org/10.1177/10534512030390020401

Purpura, D. J., & Napoli, A. R. (2015). Early numeracy and literacy: Untangling the relation between specific components. *Mathematical Thinking and Learning, 17*(2-3), 197-218. https://doi.org/10.1080/10986065.2015.1016817

Qualifications and Curriculum Authority (QCA). (2000). *Curriculum guidance for the Foundation Stage*. London: QCA, Department for Education and Employment.

Quinn, J. M., & Wagner, R. K. (2018). Using meta-analytic structural equation modeling to study developmental change in relations between language and literacy. *Child Development, 89*(6), 1956-1969. https://doi.org/10.1111/cdev.13049

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly*, *27*(1), 29-53. https://doi.org/10.2307/747832

Ramsook, K. A., Welsh, J. A., & Bierman, K. L. (2020). What you say, and how you say it: Preschoolers' growth in vocabulary and communication skills differentially predict kindergarten academic achievement and self-regulation. *Social Development*, *29*(3), 783-800. https://doi.org/10.1111/sode.12425

Reiser, R. A., Tessmer, M. A., & Phelps, P. C. (1984). Adult-child interaction in children's learning from "Sesame Street". *Educational Technology Research and Development*, *32*(4), 217-223. https://doi.org/10.1007/BF02770013

Ritchey, K. D. (2008). The building blocks of writing: Learning to write letters and spell words. *Reading and Writing*, *21*(1), 27-47. https://doi.org/10.1007/s11145-007-9063-0

Roberts, N., & Thatcher, J. (2009). Conceptualizing and testing formative constructs: Tutorial and annotated example. *ACM Sigmis Database: The Database for Advances in Information Systems*, *40*(3), 9-39. https://doi.org/10.1145/1592401.1592405

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581-592. https://doi.org/10.1093/biomet/63.3.581

Scarr, S., & Weinberg, R. A. (1978). The influence of" family background" on intellectual attainment. *American Sociological Review*, *43*(5), 674-692. https://doi.org/10.2307/2094543

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23-74. https://doi.org/10.1.1.506.1827

Sénéchal, M. (2006). Testing the home literacy model: Parent involvement in kindergarten is differentially related to grade 4 reading comprehension, fluency, spelling, and reading for pleasure. *Scientific Studies of Reading*, *10*(1), 59-87. https://doi.org/10.1207/s1532799xssr1001_4

Sénéchal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73*(2), 445-460. https://doi.org/10.1111/1467-8624.00417

Sénéchal, M., LeFevre, J. A., Smith-Chant, B. L., & Colton, K. V. (2001). On refining theoretical models of emergent literacy the role of empirical evidence. *Journal of School Psychology*, *39*(5), 439-460. https://doi.org/10.1016/S0022-4405(01)00081-4

Shamir, A., Korat, O., & Fellah, R. (2012). Promoting vocabulary, phonological awareness and concept about print among children at risk for learning disability: can e-books help?. *Reading and Writing*, *25*(1), 45-69. https://doi.org/10.1007/s11145-010-9247-x

Shonkoff, J. P., Garner, A. S., Siegel, B. S., Dobbins, M. I., Earls, M. F., McGuinn, L., Pascoe, J., Wood, D. L., & Committee on Early Childhood, Adoption, and Dependent Care. (2012). The lifelong effects of early childhood adversity and toxic stress. *Pediatrics*, *129*(1), e232-e246. https://doi.org/10.1542/peds.2011-2663

Silva, C., & Alves Martins, M. (2003). Relations between children's invented spelling and the development of phonological awareness. *Educational Psychology*, *23*(1), 3-16. https://doi.org/10.1080/01443410303218

Smith, A. K., Ayanian, J. Z., Covinsky, K. E., Landon, B. E., McCarthy, E. P., Wee, C. C., & Steinman, M. A. (2011). Conducting high-value secondary dataset analysis: An introductory guide and resources. *Journal of General Internal Medicine, 28*(8), 920-929. https://doi.org/10.1007/s11606-010-1621-5

Smith, E. (2008). Pitfalls and promises: The use of secondary data analysis in educational research. *British Journal of Educational Studies, 56*(3), 323-339. https://doi.org/10.1111/j.1467-8527.2008.00405.x

Snowling, M. J., Hulme, C., Bailey, A. M., Stothard, S. E., & Lindsay, G. (2011). *Better Communication Research Project: Language and Literacy Attainment of Pupils during Early Years and through KS2--Does Teacher Assessment at Five Provide a Valid Measure of Children's Current and Future Educational Attainments?* (Research Report No. DFE-RR172a). Cheshire: Department for Education, UK.

Stadler, M., Sailer, M., & Fischer, F. (2021). Knowledge as a formative construct: A good alpha is not always better. *New Ideas in Psychology, 60*, 100832. https://doi.org/10.1016/j.newideapsych.2020.100832

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173-180. https://doi.org/10.1207/s15327906mbr2502_4

Stephenson, K. A., Parrila, R. K., Georgiou, G. K., & Kirby, J. R. (2008). Effects of home literacy, parents' beliefs, and children's task-focused behavior on emergent literacy and word reading skills. *Scientific Studies of Reading, 12*(1), 24-50. https://doi.org/10.1080/10888430701746864

Stone, C. A., & Sobel, M. E. (1990). The robustness of estimates of total indirect effects in covariance structure models estimated by maximum. *Psychometrika, 55*(2), 337-352. https://psycnet.apa.org/doi/10.1007/BF02295291

Storch, S. A., & Whitehurst, G. J. (2001). The role of family and home in the literacy development of children from low-income backgrounds. In P. R. Britto & J. Brooks-Gunn (Eds.), *The role of family literacy environments in promoting young children's emerging literacy skills* (pp. 53-71). Hoboken, NJ: Jossey-Bass.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934-947. https://doi.org/10.1037/0012-1649.38.6.934

Strand, S., & Lindorff, A. (2021). *English as an Additional Language, Proficiency in English and rate of progression: Pupil, school and LA variation*. Cambridge: The Bell Foundation, UK.

Suizzo, M. A., & Stapleton, L. M. (2007). Home-based parental involvement in young children's education: Examining the effects of maternal education across US ethnic groups. *Educational Psychology, 27*(4), 533-556. https://doi.org/10.1080/01443410601159936

Tabors, P. O., Snow, C. E., & Dickinson, D. K. (2001). Homes and schools together: Supporting language and literacy development. In D. K. Dickinson & P. O. Tabors

(Eds.), *Beginning literacy with language: Young children learning at home and school* (pp. 313-334). Baltimore, MD: Paul H Brookes Publishing.

Taylor, C. (2018). The reliability of free school meal eligibility as a measure of socio-economic disadvantage: Evidence from the millennium cohort study in Wales. *British Journal of Educational Studies, 66*(1), 29-51. https://doi.org/10.1080/00071005.2017.1330464

Taylor, C., Joshi, H., & Wright, C. (2015). Evaluating the impact of early years educational reform in Wales to age seven: the potential use of the UK Millennium Cohort Study. *Journal of Education Policy, 30*(5), 688-712. https://doi.org/10.1080/02680939.2014.963164

Teale, W. H., & Sulzby, E. (1986). Emergent literacy as a perspective for examining how young children become writers and readers. In W. H. Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading* (pp. vii–xxv). Norwood, NJ: Ablex Publishing.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91*(4), 579-593. https://doi.org/10.1037/0022-0663.91.4.579

Townsend, M., & Konold, T. R. (2010). Measuring early literacy skills: A latent variable investigation of the phonological awareness literacy screening for preschool. *Journal of Psychoeducational Assessment, 28*(2), 115-128. https://doi.org/10.1177/0734282909336277

Uchikoshi, Y. (2019). Phonological Awareness Trajectories: Young Spanish-English and Cantonese–English Bilinguals. *Language Learning, 69*(4), 802-838. https://doi.org/10.1111/lang.12352

van Bergen, E., van Zuijen, T., Bishop, D., & de Jong, P. F. (2017). Why are home literacy environment and children's reading skills associated? What parental skills reveal. *Reading Research Quarterly, 52*(2), 147-160. https://doi.org/10.1002/rrq.160

Van Steensel, R. (2006). Relations between socio-cultural factors, the home literacy environment and children's literacy development in the first years of primary education. *Journal of Research in Reading, 29*(4), 367-382. https://doi.org/10.1111/j.1467-9817.2006.00301.x

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology*, *85*(1), 83-103. https://doi.org/10.1037/0022-0663.85.1.83

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: a 5-year longitudinal study. *Developmental Psychology*, *33*(3), 468-479. https://doi.org/10.1037/0012-1649.33.3.468

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development, 65*(2), 606-621. https://doi.org/10.1111/j.1467-8624.1994.tb00771.x

Weiss, M. J., & Hagen, R. (1988). A key to literacy: Kindergartners' awareness of the functions of print. *The Reading Teacher, 41*(6), 574-578. https://www.jstor.org/stable/20199859

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, *69*(3), 848-872. https://doi.org/10.1111/j.1467-8624.1998.tb06247.x

Whitehurst, G. J., Epstein, J. N., Angell, A. L., Payne, A. C., Crone, D. A., & Fischel, J. E. (1994). Outcomes of an emergent literacy intervention in Head Start. *Journal of Educational Psychology*, *86*(4), 542-555. https://doi.org/10.1037/0022-0663.86.4.542

Wise, J. C., Sevcik, R. A., Romski, M., & Morris, R. D. (2010). The relationship between phonological processing skills and word and nonword identification performance in children with mild intellectual disabilities. *Research in Developmental Disabilities, 31*(6), 1170-1175. https://doi.org/10.1016/j.ridd.2010.08.004

Appendix A

Table A1

Sweep 2 Main Respondent's Sex and Relationship to Cohort Member

	Natural parent	Adoptive parent	Grandparent	Total
Female	604	1	1	606
Male	6	0	0	6
Total	610	1	1	612

Table A2

Sweep 2 Partner Respondent's Sex and Relationship to Cohort Member

	Natural parent	Adoptive parent	Stepparent/partner of parent	Other non-relative	Total
Female	497	1	14	3	515
Male	6	0	0	0	6
Total	503	1	14	3	521

Appendix B

Coefficients for Pearson's Correlations Among the Observed Variables

Observed variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Covariates																					
1. Sex																					
2. Season of birth	-.01																				
3. Single-parent status	-.01	-.05																			
Family background																					
4. Maternal education	.08	-.06	.25**																		
5. Family income	.02	-.06	.53**	.47**																	
Maternal skills																					
6. Read children's books	.01	-.06	.06	.07	-.01																
7. Fill out forms	.03	-.08	.04	.18**	.07	.46**															
8. Calculate change	.05	-.07	-.04	.06	-.01	.30**	.38**														
Home tutoring																					
9. Alphabets	.00	.08	-.03	-.07	-.05	.05	.03	-.06													
10. Numbers	-.04	-.07	-.02	.01	-.04	.10*	.06	-.06	.40**												
11. Songs	.13**	-.07	-.02	.08	.03	.05	.11**	.09*	.21**	.39**											
Home literacy practices																					
12. Book reading	.03	.00	.07	.25**	.20**	.05	.15**	-.02	.20**	.15**	.18**										
13. Library visit	-.04	-.05	.13**	.18**	.13**	.01	.09*	.03	.09*	.11**	.13**	.21**									
14. Watch TV	-.07	.16**	.13**	-.25**	-.20**	-.06	-.12**	-.02	.02	.03	-.03	-.17**	-.09*								
15. Paint/draw	.12**	-.03	-.03	-.05	-.07	.06	.09*	.02	.23**	.27**	.27**	.08*	.11**	-.02							
Early skills at age 5																					
16. BAS-NV	.01	-.03	.13**	.21**	.23**	.04	.05	.02	.08	.05	.12**	.23**	.14**	-.05	.02						

17. PA	.18**	.08*	.14**	.28**	.21**	.09*	.19**	.08	.09*	.05	.12**	.19**	.10*	-.17**	.09*	.40**				
18. PC	.13**	.11**	.15**	.28**	.21**	.08	.14**	.09*	.08	.02	.07	.20**	.09*	-.08*	.06	.40**	.78**			
Later skills at age 7																				
19. BAS-WR	.15**	-.06	.18**	.31**	.23**	.05	.19**	.09*	.05	.05	.16**	.19**	.15**	-.10*	.09*	.36**	.56**	.53**			
20. Speaking/listening	.14**	.10*	.18**	.25**	.24**	.06	.14**	.14**	.11**	.03	.11**	.20**	.10*	-.01	.04	.40**	.49**	.47**	.58**		
21. Reading	.14**	.09*	.18**	.28**	.26**	.07	.19**	.12**	.08*	.04	.10*	.19**	.12**	-.05	.03	.38**	.58**	.54**	.76**	.77**	
22. Writing	.23**	.10*	.18**	.27**	.23**	.06	.17**	.09*	.07	.01	.08*	.16**	.08*	-.05	.05	.35**	.59**	.56**	.70**	.76**	.86**

Note. TV = television; BAS-NV = BAS Naming Vocabulary; PA = phonological awareness; PC = print concept; BAS-WR = BAS Word Reading. *$p < .05$. **$p < .01$.

Appendix C

Table C1

Number of Families with Income Above and Below 60% of the National Median at Sweep 1

	Frequency	Percent
Below 60% median	180	29.4
Above 60% median	432	70.6
Total	612	100

Table C2

Mother's Highest Educational Qualification at Sweep 1

		Frequency	Percent
Valid	No qualification	91	14.9
	NVQ Level 1	71	11.6
	NVQ Level 2	207	33.8
	NVQ Level 3	49	8.00
	NVQ Level 4	165	27.0
	NVQ Level 5	18	2.90
	Total	601	98.2
Missing	Overseas qualifications only	10	1.60
	Not applicable	1	0.20
Total		612	100

Table C3

Mother's Self-reported Literacy and Numeracy Skills in Daily Activities at Sweep 1

Ability	Read storybooks		Fill out forms		Calculate change	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Low	40	6.50	27	4.40	15	2.50
High	572	93.5	585	95.6	597	97.5
Total	612	100	612	100	612	100

Table C4

Frequency of Home Tutoring at Sweep 2, Age 3

	Alphabets		Numbers/counting		Songs/poems/rhymes	
Rating	Frequency	Percent	Frequency	Percent	Frequency	Percent
0	108	17.6	15	2.50	16	2.60
1	77	12.6	21	3.40	17	2.80
2	134	21.9	69	11.3	39	6.40
3	74	12.1	55	9.00	55	9.00
4	48	7.80	42	6.90	35	5.70
5	23	3.80	23	3.80	26	4.20
6	11	1.80	27	4.40	24	3.90
7	137	22.4	360	58.8	400	65.4
Total	612	100	612	100	612	100

Note. 0 = not at all, 1 = occasionally or less than once a week, 2 = 1-2 days per week, 3 = 3 times a week, 4 = 4 times a week, 5 = 5 times a week, 6 = 6 times a week, 7 = 7 times a week.

Table C5

Frequency of Reading to The Child at Home at Sweep 2, Age 3

Rating	Frequency	Percent
1	0	0.00
2	5	0.80
3	5	0.80
4	72	11.8
5	122	19.9
6	408	66.7
Total	612	100

Note. 1 = not at all, 2 = less often, 3 = once or twice a month, 4 = once or twice a week, 5 = several times a week, 6 = every day.

Table C6

Frequency of Library Visit at Sweep 2, Age 3

Rating	Frequency	Percent
0	368	60.1
1	70	11.4
2	104	17.0
3	44	7.20
4	26	4.20
Total	612	100

Note. 0 = not at all, 1 = on special occasions, 2 = once a month, 3 = once a fortnight, 4 = once a week.

Table C7

Hours a Day Child Watches Television/videos at Sweep 2, Age 3

Rating	Frequency	Percent
1	10	1.60
2	120	19.6
3	348	56.9
4	134	21.9
Total	612	100

Note. 1 = not at all, 2 = up to one hour, 3 = more than 1 hour, less than 3 hours, 4 = more than 3 hours.

Table C8

Frequency of the Child Painting/drawing at Home at Sweep 2, Age 3

Rating	Frequency	Percent
0	5	0.80
1	22	3.60
2	58	9.50
3	67	10.9
4	58	9.50
5	44	7.20
6	31	5.10
7	327	53.4
Total	612	100

Note. 0 = not at all, 1 = occasionally or less than once a week, 2 = 1-2 days per week, 3 = 3 times a week, 4 = 4 times a week, 5 = 5 times a week, 6 = 6 times a week, 7 = 7 times a week.

Table C9

Teacher-reported Number of Phonological Awareness Attributes Attained at Age 5

Attributes attained	Frequency	Percent
0	7	1.10
1	10	1.60
2	40	6.50
3	34	5.60
4	42	6.90
5	41	6.70
6	51	8.30
7	76	12.4
8	66	10.8
9	245	40.0
Total	612	100

Table C10

Teacher-reported Number of Print Concept Attributes Attained at Age 5

Attributes attained	Frequency	Percent
0	1	0.20
1	5	0.80
2	7	1.10
3	14	2.30
4	48	7.80
5	58	9.50
6	96	15.7
7	109	17.8
8	107	17.5
9	167	27.3
Total	612	100

Table C11

Teacher-reported Literacy and Language Skills at Sweep 4, Age 7

		Speaking/listening		Reading		Writing	
	Rating	Frequency	Percent	Frequency	Percent	Frequency	Percent
Valid	1	15	2.50	34	5.60	46	7.50
	2	74	12.1	106	17.3	125	20.4
	3	274	44.8	217	35.5	256	41.8
	4	182	29.7	156	25.5	125	20.4
	5	66	10.8	84	13.7	45	7.40
Missing		1	0.2	15	2.5	15	2.5
Total		612	100	612	100	612	100

Note. 1 = well below average, 2 = below average, 3 = average, 4 = above average, 5 = well above average.