Predictors of individual differences and language delay in children learning English

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Lana Susan Jago

January 2020

Declaration

This thesis is a result of my own work. No portion of this thesis has been submitted for any other degree or qualification at this University, or any other institute of learning.

Acknowledgements

I would like to express my gratitude to my supervisors, Caro and Julian. Thank you for all of your feedback and for teaching me to pursue my research interests with the best integrity. Caro, for the hours that you spent reviewing, and re-reviewing drafts; this effort has helped me become a better writer and a better researcher. I would also like to thank the rest of the members of the Language 0-5 team and everyone in the Liverpool Language Lab for their help. To my fellow "Office buddies", thank you for your endless support. I am so grateful for your friendship and your willingness to help me along the way. Charlie, this journey has been all the better for having met you. Thank you for all of your support throughout.

Thank you to all of the families who took time out of their days to come in and participate in this research. This could not be possible without their participation.

Catriona, thank you for listening to me every day over these past few years; you were a constant source of advice. Your honesty and input always helped me to keep some perspective. You are the strongest person I know.

My family, thank you for your support throughout this whole process. Mam, thank you for always encouraging me. My Siblings, for keeping things light-hearted. My sisters, Lucy and Alice, for your guidance on all things PhD and otherwise.

And to Cai, I cannot thank you enough. You have always encouraged me to do my best while providing me with unmeasurable amounts of support and love. Over the past few months, especially, thank you for your willingness to make some sacrifices and for picking up the slack, allowing me the time and freedom to bury myself in work.

This research was supported by International Centre for Language and Communication Development (LuCiD) at the University of Liverpool funded by the Economic and Social Research Council [ES/L008955/1].

Contents

Rationale for using an alternative thesis format

This thesis has been presented in the *alternative paper format*, in line with the guidelines for including research papers in a doctoral thesis provided by the University of Liverpool. This format was chosen to allow for the publication of this research in scientific journals. All three empirical chapters (chapters 2, 3 and 4) have been presented as separate manuscripts, in a structure in line with peer-reviewed journals. At the time of writing, chapter 2 is in submission in *Cognitive Development*. Chapters 3 and 4 are in preparation for submission to scientific journals. For consistency, the formatting of these papers match a common font and style used throughout the thesis. There are no reference sections provided at the end of each paper but in a single Reference section at the end of the thesis. Before each empirical chapter, there is a summary to explain how the chapters fit within the thesis and the role of each author is provided. The chapters are presented in the same format as the manuscripts that would be submitted for publication, with each chapter starting with a review of the relevant literature and ending with a discussion of the results.

Abstract

Predictors of individual differences and language delay in children learning English

Lana Susan Jago

This thesis examines individual differences in productive vocabulary development in 2-year-old children. Many factors are implicated in such individual differences in productive vocabulary development when children are 2 years old, but less is known about whether or not these factors also have a role in late talking. It is possible that the factors contributing to individual differences in vocabulary development can also be used to predict late talking. However, it is not necessarily a given that these factors will be sensitive enough to correctly identify late talking children. In addition, little is known about the predictors of language delay over time. Many children who experience an early delay in vocabulary development catch up by the time they start school, while a substantial proportion do not.

This thesis investigates these issues in detail. Chapter 1 provides an overview of the literature on individual differences in vocabulary development, on the predictors of late talking, and of later developmental delays, as indexed by parental or professional concern about development. Chapter 2, the first empirical chapter, examines the predictors of individual differences in productive vocabulary development at 24 months in a large sample of typically developing English-learning children. This chapter also establishes if those factors, implicated in individual differences, can also be used to distinguish between children who are and are not slow to learn to talk. The purpose of this was to establish if the same factors that predict individual differences can be used to categorise the speech of children with lower ability. Chapter 3 examines if a subset of these factors predicts individual differences in productive vocabulary at 25 months in a different sample of children who were identified, on the basis of vocabulary scores at 15-18 months as children likely to be very delayed in productive language (so called late-talking children), as well as typically developing children. The purpose was to determine whether those factors that successfully predict individual differences in the typical range can also be used to distinguish between typically developing and late talking children. The final empirical chapter, Chapter 4, investigates if the language measures used in chapters 2 and 3, as well as risk factors and earlier parental concern for language

development can predict concern for language development late in life, at 4-6 years of age; the age at which language disorders are starting to be diagnosed. Overall, language proficiency and cognitive measures were the most robust predictors of individual difference in productive vocabulary development at 2 years and can be used to identify children who are slow to talk and late talking children. However, language proficiency, in addition to risk factors, did not successfully identify children for whom there was concern for their language development when they were 4-6 years old. The findings of these studies and their implications for future research are discussed in the final chapter, chapter 5.

1 Chapter 1: Literature Review

1.1 Introduction to the thesis

This thesis investigates predictors of individual differences in vocabulary acquisition and late talking, and examines predictors of potential language delay, as indexed by parental or professional concern for language development. We know that there are many factors previously shown to be strong predictors of individual differences in vocabulary development (Fernald & Marchman, 2012; Kidd, Donnelly, & Christiansen, 2018). For example, previous research has shown that a number of demographic factors (Henrichs et al., 2011; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Nelson, 1973; Zubrick, Taylor, Rice, & Slegers, 2007), language proficiency (Henrichs et al., 2011; Westerlund, Berglund, & Eriksson, 2006) and cognitive ability (Fernald & Marchman, 2012; Fernald, Perfors, & Marchman, 2006) all influence the rate that children begin to produce their first words and the size of their lexicon throughout early childhood. It is plausible to think that the factors that explain why children vary in the rate they acquire language can also explain why some children experience a delay in vocabulary acquisition.

However, it is not inevitable that the factors implicated in individual differences in vocabulary development can also distinguish between children who are and are not experiencing a delay in vocabulary acquisition. It is possible that some measures can discriminate well at the top end, distinguishing between precocious children, or typically developing children, but do not discriminate well between children with lower vocabulary scores. The factors which are strong predictors of individual differences in language development may not be discriminant enough to identify children experiencing a delay in vocabulary acquisition in early childhood or children with concern for their language development later in childhood.

We address this issue in multiple ways in this thesis. The first empirical study (chapter 2) investigated, first, predictors of individual differences in vocabulary development to establish which factors contribute unique variance in productive vocabulary at 24 months after controlling for child sex and vocabulary at 18 months. These predictors include: family history of speech or language impairments, adult word count, conversational turn count, receptive vocabulary, mean length of utterances (MLU), earlier gesture use, speed of linguistic processing and non-word repetition (NWR). Following this, we established whether these predictors can correctly classify children who are and are not slow to talk (i.e. at the bottom of the typical range) in a large longitudinal, naturalistic database, in order to identify potentially promising predictors for discriminating between children with lower vocabulary scores, which can be used to identify late talking children.

The second study (chapter 3) then determined whether those factors identified as robust predictors in chapter 2, could be used to predict unique variance in productive vocabulary at 25 months in a sample that included both typically developing children and late taking children. These factors were: child sex, family history of speech or language impairments, adult word count, conversational turn count, earlier productive and receptive vocabulary, MLU, earlier gesture use, speed of linguistic processing and NWR. This extension was important to test whether, and how, the inclusion of late talking children affected the pattern of results; i.e. does the inclusion of such children affect which factors predict individual differences? Following this, we established whether we could use these predictors to distinguish between typically developing and late talking children using two classifications of late talking (1. producing fewer than 50 words at 25 months; 2. not combining words together at 25 months).

The third study (chapter 4) examined whether earlier vocabulary and gesture scores, as well as health and demographic risk factors, and parental concern, were successful at predicting potential language delays at 4-6 years, as indexed by parental or professional concern for language development. The goal here was to identify whether there was continuity across the preschool years; can we use the same predictors that successfully discriminate between children in early childhood (at 2 years of age) to identify which children are likely to have a language delay or disorder at school entry (4-6 years)?

The current chapter sets the scene for the following empirical chapters. The following three sections first summarise previous research on individual differences in language acquisition. Then evidence for the role of factors implicated in individual differences in predicting late talking is presented. Finally, a summary of the research examining the relationship between early language delays and later language impairment is provided.

1.2 Individual differences in vocabulary development

In recent years, it has become widely accepted that there are large individual differences in children's language acquisition and that these differences are crucial for predicting language development over time (Kidd & Donnelly, 2019; Kidd et al., 2018). It is in these individual differences that we begin to understand how children's current abilities influence their later learning. Among many things, children vary in the age at which they produce their first words, in their reaction time to familiar words (Fernald et al., 2006) and even in their phonological working memory capacity (Gathercole & Baddeley, 1989). In addition, children's linguistic environments vary, with some children being exposed to substantially more input than others (Hart & Risley, 1995; Golinkoff, Hoff, Rowe, Tamis-LeMonda, & Hirsh-Pasek, 2019). All of these differences influence the rate that children's language develops. While we can identify timelines for the development of many skills involved in language acquisition through the use of group testing, it is by looking at the individual differences in these skills that we begin see the trajectories for subsequent language acquisition.

Previous research has found demographic risk factors such as family income and maternal education predicts individual differences in language development (Evans, Maxwell, & Hart, 1999; Rowe, 2018; Rowe, Pan, & Ayoub, 2009). Socioeconomic status (SES; such as family income and maternal education) impacts parental input via many different factors. For example, Rowe (2018) details the impact of socioeconomic differences on parental input addressed to children; parents with higher educational attainment are more likely to provide the rich linguistic input from which children have greater opportunity to learn more vocabulary. Similarly, parents from lower socioeconomic backgrounds have limited opportunity to provide the same wealth of input compared to parents from higher socioeconomic backgrounds with more opportunity (Evans et al., 1999; Rowe, 2018; Vernon-Feagans, Garrett-Peters, Willoughby, Mills-Koonce, & The family life project key investigators, 2012). Vernon-Feagans et al. (2012) found household disorganisation predicted unique variance in children's productive and receptive vocabulary at 36 months. The research outlined here shows that there are many external factors that impact children's language development over time and how it is important to consider multiple factors when predicting vocabulary development.

Other research has shown health factors such as child sex, prematurity, low birth weight, ear infections, family history of speech or language impairment and developmental disabilities predict individual differences in language development (Harrison & McLeod, 2010; Reilly et al., 2010). For example, Reilly et al. (2010) found child sex, low birth weight and a family history of speech or language impairment predicted both productive and receptive vocabulary skills at age 4 years. Similarly, Van Noort-Van Der Spek, Franken and Weisglas-Kuperus (2012) showed that prematurity affected language development from 3-12 years, with children who were born premature having lower scores on measures of vocabulary and complex language.

There are many potential reasons why sex and family history of speech and language impairment will impact language acquisition. One possible reason is that these factors have an environmental impact on language development. Previous research has shown that parents speak more to girls than they do to boys (Cherry & Lewis, 1976; Huttenlocher et al., 1991). Two environmental factors that have been shown to play a role in individual differences in vocabulary development are caregiver input and conversational turns.

Caregiver input has been shown to be a robust predictor of individual differences in children's vocabulary acquisition (Huttenlocher et al., 1991; Jones & Rowland, 2017; Weisleder & Fernald, 2013). Children who hear more input (Hart & Risley, 1995; Hoff & Naigles, 2002) and better quality input (Cartmill et al., 2013; Jones & Rowland, 2017) have larger vocabularies. For example, Huttenlocher et al. (1991) found parental input at 16 months predicted vocabulary growth from 14-26 months. There is also research highlighting the role of conversational turn counts in children's language development in that children who engage in more conversations have larger lexicons (Romeo et al., 2018; Zimmerman et al., 2009). However, there is less research looking at the unique role of these measures in predicting vocabulary development. Typically, this research looks at the role of input or conversational turns without controlling for earlier vocabulary. One such study, by Zimmerman et al. (2009) examined the role of conversational turns in children's language over time. Conversational turns explained a unique portion of children's language scores at (on average) 28 months, after controlling for their language scores at (on average) 14 months. However, while the authors use a measure of pre-school language performance, they do not provide explicit details on the impact of conversational turns on productive vocabulary. Therefore, the unique role of both caregiver input and conversational turns requires more examination.

Children's own early language proficiency has been repeatedly shown to be a robust predictor of individual differences in later vocabulary development (Henrichs et al., 2011; Reilly et al., 2010; Rescorla, 2011; Westerlund et al., 2006). For example, children with lower vocabulary scores at 18 months were more likely to still have low vocabulary scores at 3 years (Westerlund et al., 2006). As earlier vocabulary skill is such a strong predictor of later vocabulary development, it is important to control for this when examining if other measures can predict vocabulary development over time.

When children are very young, they are producing very few words and thus it can be difficult to capture enough meaningful variance to establish relationships between early productive vocabulary and later productive vocabulary. Receptive vocabulary, therefore, often presents as a better measure of children's early communicative skills. Previous research has shown a strong relationship between children's receptive vocabulary and productive vocabulary (Jordan & Coulter, 2017) with early receptive vocabulary predicting later vocabulary development (Chiat & Roy, 2008; Stolt et al., 2016; Watt, Wetherby, & Shumway, 2006; Zambrana, Ystrom, Schjølberg, & Pons, 2013). For example, children's receptive vocabulary skills at 2;6 to 3;6 predicts variance in their productive vocabulary at 4;6 (Chiat and Roy, 2008). However, where research has predicted later productive vocabulary from earlier receptive vocabulary, typically authors tend not to control for earlier productive vocabulary. Therefore, the unique role of receptive vocabulary in language learning is still unknown.

Mean length of utterance (MLU) and the mean length of a child’s three longest utterances (M3L) are two measures of syntax, capturing the complexity of children’s first word combinations. Previous research has shown that vocabulary and syntax development are very closely related when children are young (Bates & Goodman, 1997a). For example, previous research has shown that higher vocabulary scores associated with higher MLU scores when children are between 3 and 4 years old. (Rescorla, Dahlsgaard, & Roberts, 2000; Rescorla & Schwartz, 1990). The causal relationship between syntax and vocabulary is typically reported to travel from vocabulary to syntax, with increases in vocabulary size, leading to the production of word combinations. Others have proposed that once children know a certain number of words, vocabulary and syntax emerge in parallel; children's understanding of grammar then helps vocabulary development (Dixon & Marchman, 2007; Dionne, Dale, Boivin, & Plomin, 2003). However, research looking at the relationship between syntax and vocabulary development over time, tends not to control for earlier vocabulary (Dixon & Marchman, 2007) and thus, it is still unknown if syntax plays a unique role in vocabulary development.

Early gesture skills have been previously shown to be robust predictors of vocabulary acquisition (Colonnesi, Stams, Koster, & Noom, 2010; Colonnesi et al., 2016; Rowe, Özçalişkan, & Goldin-Meadow, 2008). As with receptive vocabulary, gesture skills are an early measure of children's communicative ability and thus can be used to assess early communication in young children when vocabulary skills are still low. Furthermore, gestures have been shown to influence vocabulary development via parental input; children who produce more gestures, elicit more responses to their gestures and in turn, more parental input (Dimitrova, Özçalışkan, & Adamson, 2016; Goldin-Meadow, Goodrich, Sauer, & Iverson, 2007; Özçalışkan, Adamson, Dimitrova, & Baumann, 2017). For example, Goldin-Meadow et al. (2007) found that gestures mothers translate when children were 10-14 months were more likely to have become part of children's vocabulary by 17-23 months. Furthermore, gestures that were translated by mothers were more likely to become part of children's vocabulary compared to gestures that were not translated. However, here, and in most research that examines the role of gestures in vocabulary development, the authors do not control for concurrent vocabulary, and thus, the unique role of gestures in predicting later productive vocabulary development, above earlier productive vocabulary skills, is largely unexplored.

Speed of linguistic processing and non-word repetition (NWR) are cognitive measures that have been shown to be implicated in predicting individual differences in language development. Speed of linguistic processing measures the time it takes for children to react to familiar linguistic stimuli. It is measured by using a looking-while-listening task where children are shown two imagines on a screen and one of them is named (e.g. "look at the car"; Fernald & Marchman, 2012; Fernald et al., 2006). The time it takes for children to orient their eyes to named picture is a measure of reaction time, and has been shown to be associated with children's vocabulary (Fernald & Marchman, 2012; Killing & Bishop, 2008; Marchman, Adams, Loi, Fernald, & Feldman, 2016; Marchman & Fernald, 2008; Peter et al., 2019). For example, children's reaction times to familiar linguistic stimuli at 25 months predicts vocabulary growth from 12 to 25 months, and language and cognitive skills up to 8 years (Fernald et al., 2006; Marchman & Fernald, 2008). As with the research outlined above, where speed of processing is used to predict vocabulary development over time, there is a trend for authors not to control for earlier or concurrent vocabulary. Therefore, the unique role of speed of processing in predicting later language development is still unknown. Peter et al (2019) found that speed of processing at 19 months predicted growth in vocabulary development up until 30 months after controlling for vocabulary at 19 months. However, speed of processing at 25 and 30 months did not correlate with vocabulary from 8-37 months and as a result, the authors could not examine the role of speed of processing at these time points. Thus, the unique role of speed of processing in vocabulary development requires more research.

Non-word repetition (NWR) is a measure of phonological working memory. Children differ in their ability to temporarily store phonological representations of novel sound sequences in their working memory (Gathercole & Baddeley, 1989, 1990a, 1990b). Learning new words requires children to form long-term representations of the sound sequences associated with these words. To successfully form these long-term representations, children must first store these phonological representations in their working memory. The more speech sounds a child can store in their phonological working memory, the more opportunity there is for these words to transfer into long term memory. NWR tasks require children to repeat nonsense words back to the experimenter. NWR assesses phonological working memory via nonsense syllables; children cannot use their knowledge of the words they already know, and therefore, must store the phonological representations of the nonsense syllables in their short term memory in order to repeat them back. The more accurate children's repetitions are the bigger their phonological working memory (Gathercole & Baddeley, 1989).

Children's performance on NWR tasks is associated with their vocabulary development (Ellis & Sinclair, 1996; Hoff, Core, & Bridges, 2008; Thal, Miller, Carlson, & Vega, 2005). The majority of research examining the role of NWR in language development does so with children later in childhood (Conti-Ramsden, Botting, & Faragher, 2001; Gathercole & Baddeley, 1990b; Thal et al., 2005), and thus, little is known about the usefulness of NWR in young children at the beginning of language acquisition. One study by Hoff et al. (2008) found NWR of children aged between 20-24 months predicted variance in concurrent vocabulary after controlling for real-word repetition. However, the aim of this study was to examine if NWR was a reliable measure of phonological working memory and, thus, the authors did not evaluate the predictive power of NWR in predicting vocabulary development after controlling for earlier productive vocabulary scores. Therefore, the unique role of NWR in productive vocabulary in children this age remains unexplored.

In sum, the literature presented here demonstrates the role of these measures in language acquisition. However, in order to identify their usefulness in predicting language development over time, it is important to control for earlier vocabulary. If these measures explain no unique variance in later vocabulary skills after controlling for earlier vocabulary skills, there would be no need to use these additional measures to predict language development. In this thesis, we will first replicate and confirm some the findings presented above and then expand on that research by examining the unique role of these measure in vocabulary development. We will then examine the role of these factors in identifying children who are slow to talk and late talking children at 2 years, and then use some of these factors to predict concern for language development later in childhood.

1.3 Late talking children

As outlined above, children vary greatly in their rate of language acquisition and many factors influence language development over time. Typically, children produce their first words around their first birthday (Nelson, 1973). However, after a certain age, if a toddler has not begun to speak or is saying very few words, they will be classified as late talking. Late talking children are children who have an early delay in productive language (i.e. the number of words they can say) but are not identified as having a clinical language impairment (Rescorla, 1989, 2011; Weismer, Murray-Branch, & Miller, 1994). These children are usually identified as late talkers between the age of 18 to 35 months. Late talking children are described as producing fewer than 50 words or no word combinations at around 24 months (Rescorla, 1989; Weismer, Venker, Evans, & Moyle, 2014). This delay exists in the absence of any other developmental delays such as autistic spectrum disorder, hearing impairment or a general intellectual disability (Bishop & Edmundson, 1987; Rescorla, 2011). In addition, late talking children perform within the typical range on measures of non-verbal cognitive measures (Rescorla, 2011). Prevalence of late talking varies across the literature because different studies use different cut-off criteria. However, rates are reported to be between 10%-19% of 24 month old children (Carson & Gavin, 1998; Collisson et al., 2016; Rescorla, 1989; Rescorla & Achenbach, 2002; Zubrick et al., 2007).

Currently, we know very little about why some children experience a delay in productive vocabulary acquisition. One of the reasons why we know very little about why some children are late talkers is because much of the research on late talking has focussed on the outcomes of late talkers, establishing if their language skills catch up (Bishop et al., 2012; Dale, Price, Bishop, & Plomin, 2003; Girolametto, Wiigs, Smyth, Weitzman, & Pearce, 2001; Rescorla, 2005, 2009; Thal et al., 2005), and does not examine the predictors of late talking at the time when children are identified. Previous research, following up late talking children, has shown that they perform below their typically developing peers on measures of language development over time (Bishop et al., 2012; Fernald & Marchman, 2012; Moyle, Weismer, Evans, & Lindstrom, 2007; Rescorla, 2011; Rescorla et al., 2000; Thal et al., 2005). For example, Rescorla et al. (2000) found children who were identified as late talking between the age of 2;0 and 2;7 performed below typically developing peers on measures of syntax at 3 and 4 years. Similarly, Fernald and Marchman (2012) found late talking children, identified at 18 months, performed poorer than typically developing children on a measure of word recognition at 30 months. However, while these studies show that late talking is associated with later performance on assessments of language development, we still know very little about what causes the initial language delay in vocabulary acquisition. If we can establish predictors of late talking, we can begin to understand the underlying causes of language delay and have a better opportunity of predicting which children's language development will catch up and which children will have a persisting language impairment.

Another reason why we know very little about the causes of late talking is because much of the research on late talking focusses on distal factors. Distal factors are factors that have an indirect impact on language development. As outlined above, health and demographic risk factors impact language development, and these factors also play a role in language delay. For example, prevalence of language delay has been shown to be higher in children who experience a greater number of health related risk factors such as birth complications (Barre, Morgan, Doyle, & Anderson, 2011; Guarini et al., 2009; Jansson-Verkasalo et al., 2004), a history of early ear infections (Winskel, 2006) or a family history of speech or language impairments (Bishop et al., 2012; Collisson et al., 2016). It is also associated with biological sex, in that boys are more likely to have a language delay than girls (Reilly et al., 2010). Other research has highlighted the role of demographic risk factors on language acquisition. Low socioeconomic status (SES) is associated with increased risk for delays in vocabulary development (Evans et al., 1999; Rowe, 2018). However, these distal factors have to impact via proximal factors. To date, we still do not know what cognitive mechanisms are affected by birth complications. For example, it is possible that being premature, or of very low birth weight could impact a children's ability to efficiently process linguistic stimuli and therefore, they will be slower to learn new words. Similarly, with SES, previous research has shown that parents from lower SES families talk less to their children and produce less varied linguistic input from which children can learn (Rowe, 2018; Hoff, 2003) , and thus, SES impacts language development via input. It is, therefore, important to examine how proximal factors differ between children to establish if they can predict language development and can identify late talking children.

Individual differences in language acquisition present an opportunity to predict late talking from individual variation in children's vocabulary acquisition. As detailed above, many factors are associated with individual variation in the rate that children produce their first words. These factors that predict individual differences in productive vocabulary development may also be good predictors of late talking; individual variation encompasses slower acquisition.

In much of this thesis, we focus on four main types of predictors that previous research suggests might be promising predictors of late talking ability: sex and family history of disorders, the child's linguistic environment, the child's own communication skills at an earlier age, and the child's wider cognitive abilities. We summarise these here, but provide more detail in chapters 2 and 3. The first two - sex and family history of speech or language impairment - have been implicated in explaining individual differences in a number of studies (Bishop et al., 2012; Rescorla, 2011). Similarly, both factors play a role in language impairment. For example, Rescorla (2011) reports that incidence of late talking is more common in boys and children with a family history of speech or language impairment compared to girls. However, while being a boy or having a family history of speech or language impairment may increase a child's likelihood of being a late talker, there may be too much overlap in these factors between late talking and typically developing children, and therefore, they may not be good at discriminating between these children. That said, the inclusion of sex and family history, with other factors, may increase our ability to identify late talking children. Thus, it is important to investigate if these factors play a role in identifying late talkers.

As outlined above, two environmental factors that have been shown to predict individual differences in language development are parental input and conversational turns (Cartmill et al., 2013; Hoff & Naigles, 2002; Romeo et al., 2018; Zimmerman et al., 2009). However, there is little research looking at the role of these factors in late talking. Due to the strength of these factors in predicting variance in vocabulary development- children who hear less input and engage in fewer conversational turns have smaller vocabularies- it is likely that they can be used to identify late talking children and to distinguish between late talking and typically developing children.

Similarly, children's own communicative skills are a strong candidate for predicting late talking because they are such strong predictors of individual differences in vocabulary development. It is logical to assume that the factors implicated in delaying children's productive vocabulary development affect children's communicative development overall. Since late talking children perform at floor on measures of productive vocabulary- resulting in too little variance from which an effect can be found- measures of receptive vocabulary and gestures may be better assessments of children's communicative skills and thus are promising predictors of late talking (Duff, Nation, Plunkett, & Bishop, 2015; Hsu & Iyer, 2016; Paul & Roth, 2011; Thal, Tobias, & Morrison, 1991). Syntax and vocabulary are strongly associated in early language development (Bates & Goodman, 1997a). As mentioned above, some researchers propose that vocabulary and syntax emerge in parallel and children's understanding of grammar helps develop their vocabulary size (Dixon & Marchman, 2007; Dionne et al., 2003). Therefore, children's early syntax acquisition could be a good predictor of late talking, with delays in vocabulary development also impacting syntax acquisition. The inclusion of receptive vocabulary, gestures and syntax, in addition to measures of productive vocabulary will likely improve our ability to identify late talking children.

As outlined above, research has often focused on distal measures associated with late talking and as a result, the underlying mechanisms involved in a delay in vocabulary acquisition are unknown. Two cognitive measures that are possible candidates for predicting late talking are speed of linguistic processing and non-word repetition (NWR). Both speed of linguistic processing and NWR are associated with individual difference in productive vocabulary development (Fernald & Marchman, 2012; Fernald et al., 2006; Marchman & Fernald, 2008, Gathercole & Baddeley, 1990a). Speed of processing is a strong candidate for identifying late talking children; if children with faster processing speeds learn more words it is highly likely that children with slower processing speeds learn fewer words. Similarly, NWR is a strong candidate for identifying late talking children; if children with smaller working memory capacities have less opportunity to learn new words, they will in turn produce fewer words.

The four main types of predictors described above - sex and family history of speech or language impairment, the child's linguistic environment, the child's own communication skills at an earlier age and the child's wider cognitive abilities – are robust predictors of individual differences in language acquisition. However, it is not inevitable that the same predictors of individual differences will be the most successful predictors of late talking status. It is possible that factors are successful in predicting variation in fast learning, or all learners, but do not discriminate well between children who are and are not experiencing a delay in vocabulary acquisition. For example, while family history of speech or language impairment may be more prevalent in late talking children, most late talkers still do not have such a family history (Reilly et al., 2007). Thus, family history may not have strong discriminatory power. We aim to discover this in chapters 2 and 3. However, it is possible that combinations of measures will improve our ability to discriminate between late talking and typically developing children. Therefore, in this thesis we also examine whether combining measures that individually predict variance in early language acquisition provides a more powerful and accurate method for discriminating between late talking and typically developing children.

1.4 Linking late talking in early childhood to later language problems, as indexed by parental or professional concern for language development

An early delay in productive vocabulary development is often one of the reasons that parents first become concerned for their children's language development (Paul & Roth, 2011; Rescorla, 2011; Whitehurst & Fischel, 1994). In addition, a delay in the production of children's first words, without an clear diagnosed disability, can be an indication of a language impairment (Bishop et al., 2012). Thus, parental concern for language development can be of benefit for identifying language impairment (Glascoe, 1991; Glascoe, Altemeier, & MacLean, 1989a; Glascoe & Dworkin, 1995), and the combination of parental concern and clinical observations can improve the accuracy of paediatrician's diagnoses of developmental complications. Identifying an early delay often prompts parents to seek a referral to establish if their children's language development requires close monitoring or even intervention. Thus, parental concern for language development provides an opportunity to begin to examine delays in language development before children are old enough to be diagnosed with a language impairment. Children are typically classified as having a language impairment at approximately 5 years old (Bishop et al., 2016). However, it is evident that difficulties present earlier, and waiting for a diagnosis may potentially risk missing the opportunity to provide crucial interventions (Singleton, 2018).

Some children experiencing an early delay in vocabulary acquisition (late talkers) continue to have persisting language impairments (Bishop et al., 2012; Moyle et al., 2007; Paul & Alforde, 1993; Rice, Taylor, & Zubrick, 2008) and most perform below their typically developing peers throughout childhood (Bishop et al., 2012; Reilly et al., 2010; Rescorla, 2002, 2005, 2009). For example, Bishop et al. (2012) found that a substantial proportion of late talking (29%) children had specific language impairment (SLI, now developmental language disorder (DLD)) by the time they were 4 years old. Rice et al (2008) followed a large sample late talking children from 24 months to 7 years. Late talking children performed below the typical range on multiple measure of language ability including measures of syntax and morphosyntax. Given this, it is important to monitor children's language skills throughout childhood, once they present with an early delay in language acquisition.

However, for many late talking children, their language catches up with that of their peers in the first few years of life (Fischel, Whitehurst, Caulfield, & Debaryshe, 1989; Paul, 1996; Rescorla & Schwartz, 1990). However, for most of these children, their language skills still remain behind their typically developing peers, albeit still in the typical range, even if they are not diagnosed with a having a language impairment (Rescorla, 2002, 2005). For example, Rescorla and Schwarts (1990) followed late talking children from 22 months to 4 years. Half of the late talking children's language had caught up by age 4. However, all children, including those who had caught up, were still producing fewer syntactically complex sentences than we would expect for their age. Similarly, Rescorla (2002) found late talking children, identified between 24 and 31 months, were still behind their peers on measures of language and literacy when they were 9 years old. Children who were identified as late talkers performed significantly below their typically developing peers on measures of vocabulary, grammar, listening comprehension and reading at age 8 and were still behind on reading skills at age 9. These same children performed below their peers on language assessments up until they were 13 and 17 years old (Rescorla, 2005, 2009). Similarly, Girolametto et al. (2001) found that late talking children had, identified at 2-years -old, caught up to the typical range on grammar and vocabulary measures by the time they were 5 years old. However, these children were still performing significantly below their typically developing peers on measures of higher level language skills (e.g. pragmatic and narrative skills). The research outlined here shows that some late talkers remain delayed later in childhood and, of those that catch up to within the typical range, many still remain behind their typically developing peers. However, as of yet, there is no known method for distinguishing between children who have persisting language impairments and those children whose language will catch up. It is, therefore, crucial to establish predictors of language impairment over time.

According to the literature, many of the risk factors associated with early language delay are also associated with a later diagnosis of developmental language disorder (Campbell et al., 2003; Law, McBean, & Rush, 2011; Reilly et al., 2010). Being born premature or of low birth weight increases a child's risk of having a language impairment later in childhood (Van Noort-Van Der Spek et al., 2012). Other health risk factors such as having an ear infection also impact language acquisition; recurring early ear infection disrupts exposure to language due to a reduction in hearing and in turn impacts children's language development (Carroll & Breadmore, 2017; Winskel, 2006). Demographic risk factors impact language development throughout childhood; children from lower income families or whose mothers have lower education perform below their peers on language measures throughout childhood (Harrison & McLeod, 2010; Reilly et al., 2010).

However, while it is widely accepted that these health and demographic risk factors are associated with language acquisition much of the research on this yields only small effect sizes (Harrison & McLeod, 2010). In addition, even when a study finds a strong relationship between risk factors and later language skills, this effect disappears when examining language impairment (Reilly et al., 2010). For example, Reilly et al. (2010) found that multiple risk factors (being a boy, low birth weight, low SES and maternal education) predicted variance in productive vocabulary at 4 years. However, these risk factors were less successful in identifying children with SLI (now termed DLD); only SES and maternal education correctly classified children for whom there was a persisting language impairment.

There are many reasons why previous research examining on risk factors on language development may be inconclusive. First, it is possible that the effects of these risk factors are actually quite weak and therefore their effect is only evident in very high-powered studies using large sample sizes. For example, while Carroll and Breadmore (2017) found a history of ear infection was associated with deficits in phonological awareness, the effect was still weak. Studies with more power and more participants are often more successful at finding relationships between risk factors and language impairment over time (Kennedy et al., 2006; Van Noort-Van Der Spek et al., 2012).

Second, it is possible that risk factors are only successful predictors of language development at the group level, and do not predict on the individual level, and therefore, fail to predict language impairment. Reilly et al. (2010) found nine risk factors predicted language development up to age 4 years old. However, only three of these factors were able to accurately identify children with language impairment. It may be that predictors of individual differences in language development are not sensitive enough to discriminate between children with and without delays in language development. As mentioned above, family history of speech or language impairment is known to be implicated in language development over time. However, most children with and without language impairment do not have a family history and therefore this factor is not a good discriminator.

Third, it may be possible that risk factors are not strong enough when examined in isolation. For example, research focussing on the predictive power of risk factors, has been mixed with some research showing a significant but weak effect of some individual factors (Barre et al., 2011; Bishop et al., 2012). However, these factors may gain substantial power when combined. There is currently little research looking at how early language skills, risk factors and early parental concern for language development might interact to increase the risk of language impairment. It is possible that an early delay, combined with a greater number of risk factors, increases the likelihood of persisting language delays. For example, perhaps children who experience a delay, but have fewer risk factors, have more opportunity to catch up with their peers, where children experiencing more risk factors find the type of accelerated language acquisition needed for catch up to be more difficult. Combining risk factors could, thus, increase the likelihood of successfully distinguishing between children who are, and are not, at risk for later language disorder. Therefore, it is important to examine these early measures of language development, health and demographic risk factors, as well as early parental concern together. If a combination of these factors can predict later language problems, as indexed by later parent or practitioner concern for language development, they can be used as a starting point for monitoring language development or for informing early intervention on the basis of risk alone, before language impairment has been diagnosed.

1.5 Summary

The work here outlines the importance for identifying predictors of late talking and for examining predictors of later language delay, as indexed by concern for language impairment over time. To date, we know very little about why some children present with a delay in productive vocabulary acquisition early in childhood. Furthermore, we are still unable to distinguish between which of these children will catch up and which will continue to have persisting impairments. The aims of this thesis are 1) to better understand predictors of variance in vocabulary when children are 2 years old in order to begin to identify late talking children, and 2) to establish if some of these predictors, when combined with risk factors and early parental concern for language development at 15-18 months, can predict potential later language delays, as indexed by concern for language development when children are between 4-6 years old.

2 Chapter 2: Individual differences in productive vocabulary: Identifying children who are slow to talk.

Fit within the thesis

As discussed in chapter 1, this thesis examines unique predictors of individual differences in productive vocabulary when children are 2 years old, in order to begin to establish the role of these predictors in identifying children experiencing a delay in productive vocabulary development. The study in chapter 2 investigated whether multiple factors (child sex, family history of speech or language impairment, adult word count, conversational turn count, earlier productive vocabulary, earlier receptive vocabulary, mean length of utterances, gestures, speed of linguistic processing and non-word repetition) are unique predictors of individual differences in productive vocabulary development at 24 months. Following this, the study examined whether or not these factors can be used to identify children who are slow to talk at 24 and 30 months. The importance of this is to identify if these factors are sensitive enough to identify children with lower vocabulary scores, in order to establish if they can then subsequently be used to identify late talking children (chapter 3).

We found that multiple factors (earlier receptive vocabulary, mean length of utterances, speed of linguistic processing and non-word repetition) predicted unique variance in productive vocabulary at 24 months after controlling for child sex and earlier productive vocabulary. Furthermore, these factors, individually and combined, can discriminate between children who were slow to talk and typically developing children.

The data for this paper is part of a large longitudinal project, The Language 0-5 Project. This project ran from 2014-2019 and is part of the wider Economic and Social Research Council (ESRC) International Centre for Language and Communicative Development (LuCiD). The data was collected by Michelle Peter, Amy Bidgood and Samantha Durrant as part of the Language 0-5 Project. The study reported here was designed by Lana Jago and Caroline Rowland. Lana Jago was responsible for all coding of the data, for all analyses and for writing the paper. Lana Jago and Caroline Rowland worked in collaboration to construct the final manuscript and all authors provided comments on this manuscript. At the time of submitting this thesis, this chapter has been submitted as a paper to the journal *Cognitive Development*.

Abstract

This study investigated why some children are slower than others at acquiring vocabulary, using a new, unique intensive longitudinal sample of 79 English-learning children, tested every 2-3 months between 8 and 30 months of age (the Language 0-5 Project children). First, we investigated what are the best predictors of individual differences in productive vocabulary development at 24 months. Next we assessed the discriminatory ability of these predictors in identifying children who are slow to learn to talk at 24 and 30 months old. Regression analyses revealed conversational turn counts, earlier receptive vocabulary, mean length of utterance (MLU), non-word repetition (NWR) and speed of linguistic processing were all predictors of individual differences in productive vocabulary when controlling for earlier productive vocabulary and sex. Receiver operation characteristic curve and discriminant function analyses revealed that these predictors were also successful at distinguishing between children who are, and are not, slow to talk.

2.1 Introduction

A substantial proportion of children are slow to learn to talk. These children, whose productive language development is delayed compared to their peers, exhibit no other measurable cognitive delays, and have experienced no hearing or visual impairment that can explain this delay in language acquisition (Bishop & Edmundson, 1987; Rescorla, 1989).

In fact, we know very little about why some children are slow to acquire language in the first two years. This lack of understanding is problematic for two reasons. First, on a theoretical level, determining the causes of slow development will help solve the puzzle of individual differences; it will help us understand why children differ so substantially in their rate of language acquisition in the first two years of life. For example, if we find that slow language learners are slower to process sentences (Fernald & Marchman, 2012) or have smaller phonological working memories (Thal, Miller, Carlson, & Vega, 2005), this will provide important evidence for the role of these processes in early language acquisition.

Second, having a better understanding of why some children are slow to learn to talk has practical implications, since some late developers go on to develop developmental language disorder (DLD, Bishop & Adams, 1990; Rescorla, 2002), though precise proportions vary depending on the age that persisting language delay is identified and the criteria used to characterise it (Law, Boyle, Harris, Harkness, & Nye, 2000). For example, Rescorla and Schwartz (1990) found that 54% of children identified with an early expressive delay at 2 years were still delayed on measures of productive language skills at a follow up visit between 3-4 years old. In a later study, following up the same children, Rescorla (2002) found that 17% of these children had DLD at 6 years old and many were still delayed on multiple language measures when they were 9 years old. Therefore, the more we understand about the various causes of slower language learning early in acquisition, the better chance we have of predicting which children might go on to develop language delay.

Unfortunately, one of the reasons why we know very little about why some children are slow to talk is because the bulk of the research on this topic focusses on attempts to determine which of these children will go on to develop a developmental language delay, rather than the factors that predict late talking itself (see e.g. Dale, Price; Bishop & Plomin, 2003; Rescorla, 2002; 2005; Rescorla & Schwartz, 1990).

Children who experience an early delay in vocabulary development are at a greater risk for developing persisting language impairments. In these studies, late talkers are identified on productive vocabulary alone, leaving us with little understanding about other processes involved in early vocabulary acquisition. Without understanding the mechanisms involved in early language acquisition, we cannot begin to determine why some children experience a delay in productive vocabulary.

In addition, what work there is on predicting why some children are slow to talk tends to focus on distal factors (factors that only have an indirect effect on language development; e.g. socioeconomic status (SES), birth complications, sex) rather than proximal factors (direct factors such as input quantity and quality; Zubric, Taylor, Rice & Slegers, 2007; Rescorla, 1989; Reilly et al., 2007). Children who are exposed to more distal risk factors, such as low SES or low birth weight, are more likely to be slow to develop productive vocabulary (Reilly et al., 2007). Since distal factors have to have an influence via proximal factors, it is important to investigate what proximal factors influence late talking. For example, we know that birth complications affect language development, in that the prevalence of slow language development is higher in premature children (Jansson-Verkasalo et al., 2004). However, we do not know which cognitive mechanisms are affected by prematurity; for example, it could be that prematurity affects children's ability to process linguistic material quickly, which explains why they are slow to learn vocabulary. Similarly, the relationship between SES and late talking may be explained by the fact that parents from lower SES families tend to talk less to their children (Hoff, 2003; Hoff-Ginsberg, 1991). Thus, in the present study, we focus on how proximal factors affect language development. If we can identify which proximal factors are implicated in early language acquisition, we can begin to understand how they contribute to delays in vocabulary development.

We can identify a range of promising predictors of why some children are slower than their peers by looking at the larger literature on individual differences in vocabulary development within the typical range. Although most children begin to comprehend words within their first year and produce words soon after their first birthday, there is much individual variation in the timing of language acquisition trajectories, with more precocious speakers starting earlier, and reaching language milestones faster, than their peers (Fenson et al. 1994; Fernald & Marchman, 2012; Kidd, Donnelly, & Christiansen, 2018). For example, data from the norming sample of the MacArthur-Bates CDI (see Stanford Wordbank database ,http://wordbank.stanford.edu) shows that the top 10th percentile of children at 24 months are, on average, producing 549 words, compared to the 84.6 words produced by the bottom 10th percentile.

We already know quite a lot about what predicts individual differences in vocabulary development (e.g. demographic factors such as sex and family history of speech or language impairment, environmental factors such as input and conversational turn taking). Many of these factors are also candidates for predicting which children may be slow to learn to talk, since it seems logical to assume that the same factors that predict why some children are faster than average are likely to also explain why some children are slower. In addition, many of these factors are implicated in individual differences during the time children first exhibit delays in productive vocabulary development - in the first two years of life. Therefore, we know that these are integrally implicated in the early stages of language acquisition.

However, it is not inevitable that predictors of individual differences are the same as predictors of slow talking. Some measures may discriminate well between faster learners, or between learners in general, but have very little discriminatory power when distinguishing between children with lower vocabulary scores (e.g. boys tend to be slower than girls on average but the differences are so small, and the overlap in standard deviation so large, that sex has very little discriminatory power). Thus, it is important to investigate whether the predictors implicated in individual differences across the spectrum can also be used to discriminate at the lower end of the scale. The goal of the present study was to investigate whether 10 factors, all implicated in explaining individual differences in vocabulary size during the first two years of life, can also be used to discriminate between children whose productive vocabulary is of lower ability: child sex, family history of speech or language impairment, input quantity, conversational turn taking, productive and receptive vocabulary size at earlier time points, language complexity (MLU), gesture use, speed of language processing and phonological working memory.

We used data from a new unique longitudinal cohort dataset (the Language 0-5 project), which ran in the North West of England and followed 95 children from 6 months to 4 years and 6 months. Data on a large range of factors associated with early language development were collected at multiple time points each year, which allowed us to assess the influence of a number of factors at once, and to establish if each contributes unique variance to vocabulary acquisition. It is important to note that few, if any, of the Language 0-5 children have language that is so delayed that it warrants concern. However, there are large individual differences in productive vocabulary size within the sample, which we can use to identify characteristics of children who are slow to learn to talk (henceforth slow-to-talk children), and then apply in further studies to determine if they also allow us to identify children whose language is worryingly delayed.

In the remainder of this introduction, we describe why each of the predictors mentioned above are promising candidates for identifying children whose productive vocabulary is of lower ability, before summarising the objectives of the study.

2.1.1 Demographic factors

Two demographic factors robustly implicated in explaining individual differences in vocabulary acquisition are child sex and family history of speech or language impairment. Sex emerges as a robust predictor of individual differences in many studies, with boys generally tending to develop more slowly than girls (e.g. Acredolo & Goodwyn, 1988; Henrichs et al., 2011; Horwitz et al., 2003; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Nelson, 1973). In addition, boys tend to be strongly represented at the lower end of the scale (Horwitz et al., 2003, Nelson, 1973; Rescorla, 1989). Thus, sex is a likely predictor of being slow to talk, as well as individual differences in vocabulary acquisition, although, as we explain above, we think it unlikely to have much discriminatory power.

Family history of speech or language impairment is also a strong candidate for discriminating between children whose productive vocabulary is of lower ability, since it has been consistently reported to be predictive of individual differences in early vocabulary development (Bishop et al., 2012; Collisson et al., 2016; Hadley & Holt, 2006; Reilly et al., 2007; Reilly et al., 2009; Reilly et al., 2010; Zubrick, Taylor, Rice & Slegers, 2007), with a positive family history of speech or language impairment being strongly associated with poorer productive vocabulary skills (Bishop et al., 2012; Olswang, Rodriguez, & Timler, 1998; Rescorla, 2011; Zubrick et al., 2007). While there is some research looking at the relationship between family history of speech or language impairment and late talking, there is often a failure to control for earlier productive vocabulary or sex, making it difficult to determine whether family history is a robust unique predictor. Therefore, the unique role of family history in predicting whether children will be slow to learn to talk is still unexplored.

2.1.2 Environmental measures

There are two potential reasons why sex and family history of speech or language impairment, discussed above, might affect language acquisition. First, it is highly likely that there is a biological component; boys may mature more slowly than girls, and children with a family history of speech or language impairment are at an increased risk of having inherited a speech and language disorder from their parents (Zubrick et al., 2007; Reilly et al., 2010; Rescorla, 2011). However, both factors may also have an environmental component; parents may talk more to girls (Cherry & Lewis, 1976; Huttenlocher et al, 1991) and parents with speech and language difficulties themselves may find it harder to model rich linguistic input when talking with their children. Two such environmental factors, in particular, are robustly implicated in explaining individual differences in vocabulary acquisition: caregiver input, and number of conversational turns.

Caregiver input plays an important role in the individual differences children exhibit in vocabulary development (Hirsh-Pasek et al., 2015; Weisleder & Fernald, 2013; Huttenlocher et al., 1991), with a wealth of evidence suggesting that children who hear more input, and better quality input, start to talk earlier, and acquire language faster (Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010; Huttenlocher et al, 1991; Hart & Risley, 1992; Pearson, Fernandez, Lewedeg, & Oller, 1997). Similarly, there is evidence to suggest that children who engage in more conversational turns perform better on measures of vocabulary development (Zimmerman et al., 2009; Romeo et al., 2018). While there is little in the literature that directly assesses the role of input and conversational turns in predicting late talking, the fact that they are such strong predictors of individual differences makes them likely candidates. Furthermore, we know that input and conversational turns are affected by SES and that the incidence of late talking is more prevalent in low SES families (Hirsh-Pasek et al., 2015; Romeo et al., 2018; Rowe, Özçalişkan & Goldin-Meadow, 2008; Weisleder & Fernald, 2013). Thus, SES may be able to distinguish between children whose productive vocabulary is and is not of lower ability, in part because children from low SES families hear fewer utterances and engage in fewer conversations compared to children from high SES families.

2.1.3 Language Proficiency Measures

A highly robust predictor of individual differences in vocabulary at 24 months is productive vocabulary at an earlier time point (Fernald & Marchman, 2012; Henrichs et al., 2011; Rescorla, 2011; Westerlund, Berglund, & Eriksson, 2006). For example, Henrichs et al. (2011) found vocabulary scores at 18 months explained 11.5% of the variance in vocabulary scores at 30 months. Therefore, productive vocabulary at earlier time points is also likely to be a good predictor of whether a child will be slow to learn to talk.

However, it can be difficult to find clear relationships across time between productive vocabulary scores early in life, simply because we find floor effects since very young children do not produce many words. Thus, size of receptive vocabulary may actually be a stronger predictor of later productive vocabulary. In fact, earlier receptive vocabulary skills have been shown to predict individual differences in later productive vocabulary skills in a number of studies (Stolt et al., 2016; Watt, Wetherby, & Shumway, 2006; Zambrana, Ystrom, Schjølberg, & Pons, 2013) with lower receptive vocabulary scores associated with lower productive vocabulary scores (Hsu & Iyer, 2016; Paul & Roth, 2011). Therefore, earlier receptive vocabulary is likely to be good at distinguishing between children with lower productive vocabulary scores. While there is some research looking at the role of receptive vocabulary in identifying children who are slow to talk (Thal, Bates, Goodman & Jahn-Samilo, 1997; Paul, Looney, & Dahm, 1991), this research often fails to control for earlier productive vocabulary skills, meaning the unique role of receptive vocabulary in predicting delays in productive vocabulary acquisition remains largely unexplored.

Another early learned communicative skill which shows large individual differences early in life is gesture production (Colonnesi, Stams, Koster, & Noom, 2010; Hsu & Iyer, 2016; Iverson & Goldin-Meadow, 2005; Özçalişkan & Goldin-Meadow, 2005; Rowe et al., 2008; Rowe & Goldin-Meadow, 2009; Schults, Tulviste, & Konstabel, 2012; Thal, Tobias & Morrison, 1991; Zambrana et al., 2013). Children elicit parental input through gestures and in turn learn from this input; when parents translate gestures into words, these words are more likely to enter a child's vocabulary compared to the words for gestures which parents do not translate (Goldin-Meadow, Goodrich, Sauer & Iverson, 2007; Özçalışkan, Adamson, Dimitrova, & Baumann, 2017; Dimitrova, Özçalışkan, & Adamson, 2016). Gesture use might be predictive of delays in productive vocabulary development for two reasons. First, if children are slow to develop the ability to communicate, this might affect both verbal and non-verbal communicative abilities. Second, since gestures are a way for pre-verbal children to initiate communicative exchanges with adults, parents may interact less with, and thus address less language to children who gesture infrequently (Dimitrova et al., 2016), which in turn will reduce the amount of input from which children can learn new words.

Finally in this section, mean length of utterance (MLU) is a measure of language complexity that assesses how many morphemes (or sometimes, words) there are, on average, in children's sentences at different stages of development (Brown, 1973). Unsurprisingly, individual differences in MLU are associated with individual differences in vocabulary size (Bates & Goodman, 1997; Rescorla, Dahlsgaard, & Roberts, 2000; Rescorla & Schwartz, 1990). The causal relationship between MLU and vocabulary development is likely to be from vocabulary to MLU (Mc Gregor, Sheng and Smith, 2005; Petinou & Spanoudis, 2014; Smith, & Jackins, 2014). However, Dixon and Marchman (2007) suggest that vocabulary and grammar emerge in parallel; children's grammatical complexity and productive vocabulary both increase with age. Similarly, Dionne, Dale, Boivin, and Plomin (2003) found grammatical patterns were abstracted from the lexicon and that this acquired grammatical knowledge, in turn, aids further vocabulary acquisition. Therefore, it is worth investigating whether MLU adds predictive power when attempting to distinguish between children whose productive vocabulary is of lower ability.

2.1.4 Cognitive measures

Two cognitive measures that have been shown to predict individual differences in vocabulary development are speed of linguistic processing and non-word repetition (NWR). Speed of linguistic processing is a reaction time measure and refers to the speed with which a child is able to process familiar linguistic material. It is measured by testing how quickly children orient their eyes to a target picture on hearing its label (e.g. "look at the ball": Fernald et al, 2006). The rate at which children can recognise familiar words has been shown to be related to their vocabulary development (Fernald & Marchman, 2012; Fernald, Perfors & Marchman, 2006; Killing & Bishop, 2008; Marchman, Adams, Loi, Fernald, &

Feldman, 2016;). For example, Fernald et al. (2006), found speech processing at 25 months was related to individual differences in the trajectories of vocabulary development from 12-25 months; children whose vocabulary developed at a more accelerated rate during this time performed better on the speech processing task. The precise mechanism through which this happens is not known; it may be that faster processing of familiar words frees up resources that can be dedicated to learning new words (Fernald & Marchman, 2012) or that having a larger lexical network improves lexical processing speed because these children are likely to have had more experience hearing and using speech (Fernald et al., 2006). Either way, if children with faster processing speed have larger vocabularies, it seems highly likely that children who process language slowly will learn language more slowly, Thus, speed of processing seems like it could be a promising candidate for identifying children who are slow to talk.

Finally, individual differences in children's vocabulary development are also associated with individual differences in their ability to store phonological representations of novel sound sequences in their working memory (Ellis & Sinclair, 1996; Gathercole & Baddeley, 1990a). In order to learn new words, children must form a long-term representation of the sound sequences associated with each word. To form this representation, children need to first store this representation, temporarily, in phonological working memory. Therefore, the more information a child can store in their phonological working memory, the greater the opportunity for this information to be processed and stored in long-term memory (Gathercole & Baddeley, 1990a).

Differences in phonological working memory are measured using non-word repetition (NWR) tests in which children are asked to repeat nonsense words back to an experimenter. Performance on NWR tasks is known to be robustly associated with individual differences in vocabulary development (Duff, Nation, Plunkett, & Bishop, 2015; Hoff, Core & Bridges, 2008; Petinou & Spanoudis, 2014; Stokes & Klee, 2009; Thal et al., 2005). NWR is used to identify older children with language disorders, with poorer performance on NWR tasks associated with lower scores on language assessment (Botting & Conti-Ramsden, 2001) and language impairment (Conti-Ramsden, Botting & Faragher, 2001). Few studies have tested the relationship between NWR performance and vocabulary in preschool children under 4 years, but those that do report a robust relationship (Hoff et al, 2008; Roy & Chiat, 2004; Torrington Eaton, Newman, Ratner & Rowe, 2015). We predict that, because low scores on NWR tests are predictive of poorer productive vocabulary, NWR is likely to be a strong candidate for distinguishing between children whose productive vocabulary is of lower ability.

In summary, the goal of the present study was to investigate whether the 10 factors described above, all implicated in explaining individual differences in productive vocabulary size in the first two years of life, could also be used to classify children who are slow to talk, in a cohort of 79 children aged 24 and 30 months. Identifying which measures are sensitive enough to identify children whose productive vocabulary is of lower will allow us to begin to use these measures for identifying late talking children in future studies. We first used correlations and regressions to determine which of these factors predict individual differences in productive vocabulary size at 24 months in our sample. Then, we used sensitivity and specificity analyses and discriminant function analyses to establish the discriminatory ability of these factors in classifying children as slow-to-talk at 24 and, 6 months later, at 30 months of age. The participants for this study were part of a longitudinal project run in the North West of England, the Language 0-5 Project. The data for this study was collected from multiple data points from when the children were 6 months old up until they were 30 months old.

2.2 Methods

2.2.1 Participants

An initial sample of 89 families were recruited at 6 months of age, at the beginning of the project in 2014. An additional six families were recruited at the 15 month data point to replace some who had dropped out. A total of 95 families were recruited. One family was excluded due to their responses on a family background questionnaire completed by the parents at 6 months (persistent ear infections likely to affect hearing). At the time of writing, 15 families (16%) in total had dropped out of the study due to time commitment. Two families had dropped out after the initial sign-up and a further three after the first visit at 6 months. Five families dropped out after the 8 months data-point and one after each of the 14, 15 and 19 month data-points. A final two dropped out after the 21 month data-point.

At the time of recruitment, at 6 months of age, all infants were born full-term, none were born low birth weight and all were typically-developing. We had available 24 month productive vocabulary scores for 75 children, and 25 month productive vocabulary score for a further four, which we used as a proxy for their 24 month data. We had available 30 month productive vocabulary scores for 73 children. Thus, there were 79 participants in our full sample at 24 months and 73 in our full sample at 30 months, though there were fewer for some analyses because we did not have full samples of predictor variables for all children. Table 2.1 describes the sample characterises at each data point.

The sample was split almost evenly for sex (girls =41). More than half the children were first borns (n=50) with the remainder having one or more older siblings (1 older sibling: n = 24; 2 or more older siblings: n = 5). All children were mono-lingual English learners. The children were specifically recruited to have no developmental disorders that would influence language development. Therefore, no children suffered from persisting ear infections (e.g. glue ear).

Table 2.1

Number of participants, sex, mean age and age range for each variable at each data point

Category	Variable	Girls				Boys			
		n (%)	Data point (months old)	Mean Age (y;m;d)	Age Range (y;m;d)	*n* (%)	Data point (months old)	Mean Age (y;m;d)	Age Range (y;m;d)
Demographic	Sex	41 (52%)	At intake	0;7;11	0;5;24-1;3;21	38(48%)	At intake	0;6;27	0;5;29-1;3;16
	Family history	41 (52%)	At intake	0;7;11	1;6;5-1;7;21	38(48%)	At intake	0;6;27	1;6;6-1;7;18
Input variables	Conversational turn count at 18 months	40(51%)	18	1;6;24	1;6;5-1;7;21	38(49%)	18	1;6;26	1;6;6-1;7;18
	Conversational turn count at 21 months	41(52%)	21	1;9;12	1;9;0-1;9;25	38(48%)	21	1;9;12	1;8;28-1;9;28
	Adult word count at 18 months	40(51%)	18	1;6;24	1;6;5-1;7;21	38(49%)	18	1;6;26	1;6;6-1;7;18
	Adult word count at 21 months	41(52%)	21	1;9;12	1;9;0-1;9;25	38(48%)	21	1;9;12	1;8;28-1;9;28
Language variables (speech):	18 month receptive vocabulary	40(51%)	18	1;6;16	1;6;0-1;7;1	38(49%)	18	1;6;19	1;6;4-1;7;3
	18 month productive vocabulary	40(51%)	18	1;6;16	1;6;1-1;7;2	38(48%)	18	1;6;19	1;6;4-1;7;3
	19 month M3L	37(47%)	19	1;7;11	1;6;26-1;7;28	35(45%)	19	1;7;12	1;7;0-1;8;0
	24 month productive vocabulary	41(53%)	24	2;0;14	1;11;20-1;1;15	38(49%)	24	2;0;15	1;11;19-2;1;18
	24 month MLU	40(56%)	24	2;0;13	1;11;24-2;1;0	37(51%)	24	2;0;12	1;11;22-2;1;10
Gesture use	8 month gestures	36(46%)	8	0;8;6	0;8;0-0;8;19	36(46%)	8	0;8;5	0;7;27-0;8;20
	9 month gestures	31(40%)	9	0;9;22	0;9;0-0;10;18	33(43%)	9	0;9;20	0;9;8-0;10;6
	11 month gestures	34(47%)	11	0;11;9	0;11;0-0;11;25	33(46%)	11	0;11;10	0;10;29-1;0;2
	12 month gestures	37(58%)	12	1;0;7	0;11;28-1;0;22	35(55%)	12	1;0;11	0;11;22-1;1;15
	15 month gestures	41(61%)	15	1;3;10	1;3;0-1;3;24	36(54%)	15	1;3;11	1;3;0-1;4;2
	16 month gestures	39(54%)	16	1;4;27	1;4;10-1;5;21	36(50%)	16	1;4;29	1;4;4-1;5;25
	18 month gestures	40(52%)	18	1;6;16	1;6;0-1;7;1	38(49%)	18	1;6;19	1;6;4-1;7;3
Cognitive variables	19 month speed of linguistic processing	39(52%)	19	1;7;12	1;7;0-1;7;27	37(49%)	19	1;7;13	1;7;0-1;8;0
	25 month non-word repetition	35(45%)	25	2;1;12	2;1;0-2;1;28	31(40%)	25	2;1;16	1;0;24-2;2;10

2.2.2 Overall design

This was a longitudinal study using experimental, questionnaire and observational methods. This study was granted ethical approval by The University of Liverpool's Research Ethics Subcommittee for Non-Invasive Procedures for the study Language Development in Late Talkers (Institute Review Board (IRB) protocol number: RETH000764). We ran two different sets of analyses. In the first set of analyses we used correlational and regression designs to test which of our 10 factors predicted individual differences in productive vocabulary size at 24 months. For the regressions, the independent variables were: demographic factors (sex and family history of language delay or dyslexia); environmental factors (conversational turn count and adult word count), language proficiency measures (productive and receptive vocabulary scores at 18 months, MLU at two time points (19 and 24 months)), gesture scores at 7 time points (8, 9, 11, 12, 15, 16 and 18 months), and two measures of cognitive ability (speed of linguistic processing and non-word repetition performance, which measures phonological working memory). The dependent variable was productive vocabulary scores at 24 months as measured by the Lincoln Communicative Development Inventory (a UK adaptation of the MacArthur Bates CDI Words and Sentences; Meints, Fletcher, & Just, 2017).

The second set of analyses used receiver operating characteristic (ROC) curves and discriminant function analyses to establish the discriminatory ability of the factors in classifying children into language groups at 24 months and 30 months. We chose 24 months, because this is the age at which late talkers are most often identified in the literature. However, since the factors chosen for these analyses were determined on the basis of regressions performed on the 24 month data, we ran the final set of analyses – the discriminant function analyses – again, this time on the productive vocabulary of the children at 30 months, to verify the pattern of results on a different, but related, outcome measure. The independent variables were the six predictors that emerged from analysis set 1 as significant predictors of individual differences in our sample: sex, earlier productive and receptive vocabulary scores, MLU at 24 months, speed of linguistic processing and non-word repetition. The dependent variables were language group at 24 months for the ROC curve analyses and language group at 24 and 30 months for the discriminant function analyses.

The data collected for the Language 0-5 Project includes data from multiple time points (see Rowland, Bidgood, Durrant, Peter & Pine (unpub.), available on the Open Science Framework at https://osf.io/kau5f/) so we wanted to make hypothesis-based, *a priori,* decisions about which data points to use for each of our predictors in order to remove the potential for p-hacking. Since our dependent variable was productive vocabulary at 24 and 30 months, where possible our predictor variables were based on data collected at our 18 and 19 month data points, which were the first major data points prior to 24 months that included measures that have been used in previous research to predict late talking, as well as being the age most often tested in the literature on late talking prior to 24 months.

However, there were four exceptions to this. First, for gesture use, we used data from multiple data points (7 data points between 8 and 18 months) since it is not clear from the literature when we might expect this effect to emerge. Second, for the adult input measures, we used data averaged across two data points (at 18 and 21 months of age), in order to improve the representativeness of the data (i.e. to increase the chances that the data is representative of the child's typical daily environment). Third, phonological working memory data was not collected before 25 months, since children younger than this do not have enough productive language to complete the test successfully (Eaton, Newman, Ratner, & Rowe, 2015; Gathercole & Adams, 1993). Thus, we used data from the earliest data point we could: the 25 month data point. Fourth and finally, we supplemented the child MLU measure we obtained at 19 months (M3L from the Lincoln CDI) with a measure of MLU calculated from transcripts of natural conversations between the child and a caregiver at the 24 month data point. We did this because we wanted to include a traditional measure of MLU (calculated from spontaneous speech), as well as the M3L measure from the CDI, but did not have transcripts available to calculate this at 18 or 19 months.

2.2.3 Demographic measures

Here we used the Family Questionnaire and the Family Language History Questionnaire to ascertain sex and family history of language delay or dyslexia. The Family Questionnaire is a simple factual questionnaire that asks a range of questions about a child's health and family life. It was devised for the UK-CDI project (for details of construction, see Alcock, Meints, Rowland, Brelsford, Christopher, & Just, 2020). It includes a question about the child's sex, and a question about family history of language delay or dyslexia (the questions of interest were: “4. Is there anyone in the child’s immediate family (brothers/sisters/parents only) with a speech or language difficulty or dyslexia?” (yes/no)). Parents filled this in at intake, then more details were elicited using a more specific Family Language History Questionnaire (Alcock et al., 2020) completed when the children were 18 months. This questionnaire asks parents about the language history of each of the child’s immediate family members (e.g. mother, father, brother, half-sister). Questions include details on each person’s speech and language development (e.g. “Was she/he identified as being a late talker?”). Nine of the children had a family history of language delay or dyslexia.

2.2.4 Environmental measures

The Language Environment Analysis (LENA; LENA Research Foundation, 2014) system was used to record audio from the children’s environment. Here we use the data collected at both the 18 and 21 month data points, since we wanted to predict later vocabulary from earlier input (Hirsh-Pasek et al., 2015; Weisleder & Fernald, 2013). Parents were given a LENA Pro digital language processor (DLP; LENA Research Foundation, 2014) for their child to wear in a specialised t-shirt with a small padded pocket on the front. Parents were asked to record a full 16 hours during one typical day at both 18 and 21 months. Actual hours recorded ranged from 8 hours to 16 hours (M=15 hours, 14 minutes). We used two measures to assess language in children’s environment: adult word count (input) and conversational turn count. The LENA Pro processing software (LENA Research Foundation, 2014) automatically analyses the audio recording and provides estimates of adult word count and conversational turn count. Previous research has found LENA calculations to correlate highly with human coding (Gilkerson et al., 2017; Zimmerman et al., 2009). Adult word count is the total number of words spoken by adults within the child’s environment. One limitation of the measure of adult word count captured using the LENA system is that it collects all adult speech and is not limited to child directed speech. We revisit this in the discussion section. A conversational turn is calculated when a child utterance is followed by an adult utterance, or vice versa, within a 5 second window. Total adult word count and conversational turn count were averaged from both time points to provide one value for each measure. For one child, we only had 21 month data available so we used this data as their calculated adult word count and conversational turn count. The mean adult word count was 18,238.25 with a wide range, indicating that there were substantial differences across our families (range=7,239.00 - 43,926.50). The mean number of conversational turns was 806.11 (again, with substantial individual difference: range=300.50 - 1804.00).

2.2.5 Language proficiency measures

2.2.5.1 CDI questionnaires

Communicative Development Inventories (CDIs) are widely used parent report checklists of words, and gestures or sentences, which require parents simply to indicate if their child uses, or understands, an item. The UK-CDI Words and Gestures (Alcock, et al., 2020) was used to collect productive and receptive vocabulary, and gesture scores from 8-18 months (total possible scores are 396 for vocabulary and 57 for gesture). The UK-CDI has been standardised for the UK population, and has good validity and reliability (see Alcock et al., 2020). The Lincoln CDI was used to collect productive vocabulary scores at 24 and months (total possible score = 689) and MLU scores (M3L; no upper bound) at 19 months. The Lincoln CDI is the UK version of the MacArthur Bates CDI Words and Sentences (Meints, et al., 2017). It has not yet been validated but is expected to have good validity and reliability due to its similarity to the MacArthur-Bates instrument. For both instruments, scores were calculated according to the instructions in the manuals referenced above.

2.2.5.2 Lab-based naturalistic play

At the 24 month data point, children were video-recorded for a 30 minute naturalistic play session in the lab, based on the procedure of Quinn, Donnelly, and Kidd (2018). The play sessions were split into three ten minute sessions that ran back-to-back. Children played with one set of toys for the first ten minutes and another set for the second ten minutes. For the last ten minutes, children played with both sets of toys. One set of toys was chosen to elicit functional play. The other set of toys was chosen to elicit symbolic play. We counterbalanced which set of toys children played with in the first two ten minute sessions. One parent or caregiver was involved in the play session, typically the mother. Parents were told that the session should be child-led; allowing their child to pick the toys.

Children's speech from the recordings was transcribed to provide our 24 month MLU measure. These recordings were transcribed in CHAT format and morphologically coded using the MOR program (Macwhinney, 2000), then MLU for each child was calculated as number of morphemes per utterance, using the associated CLAN software. Utterances with self-repetition, imitations or routine speech were excluded from our analysis. Three MLU scores were calculated for each 10 minute session and then a mean MLU was calculated across the 30 minutes. There were substantial individual differences in MLU across children (MLUs ranged from 0.89 to 3.30; mean = 1.79). Finally, to check that the speech recorded in the lab was representative of the child's speech at home, we transcribed naturalistic data collected in the home for seven children and ran correlations between MLU taken from lab and home recordings. Reliability was extremely good (r = .90) so we can be confident that MLU calculated from the lab-based recording reflects children's linguistic competence.

2.2.6 Cognitive measures

Speed of processing

A looking-while-listening paradigm using an EyeLink ® 1000 Plus eye-tracker (EyeLink 1000 Plus, SR Research Ltd) was used to capture children's speed of linguistic processing at the 19 month data point. Children heard a series of 64 sentences paired with visual stimuli, across two blocks (A and B). Two pictures were presented on the screen paired with a pre-recorded sentence naming one of the pictures (e.g., Where's the baby? Can you see it?). Words which appeared as the target in block A, appeared as the distractor in block B. Each block also included three filler sentences (e.g., "Here comes the train! Choo-choo"), to keep the child interested. Children were randomly assigned to a counterbalance group ("A then B" or "B then A"). Prior to the session, parents indicated on a questionnaire whether or not their child understood the target words.

Speed of processing was measured in terms of reaction time (RT). RT was defined as the time taken to initiate a shift from the distractor picture to the target picture between 300ms after and up to 1800ms target word onset. Trials where the child was not looking at the distractor image at the onset of the target word were excluded from the analysis. RT was calculated per trial, per child. A total mean RT for the task was calculated for each child and used for our analyses. For the full details of stimuli creation, procedure and coding, see Peter et al. (2019).

Non-word repetition

The NWR task used here was designed by the Language 0-5 team and was completed at the 25 month data point. Full details of stimuli creation, procedure and coding can be found in The Language 0-5 Project Non-Word Repetition Test (Bidgood, Durrant, Peter, Pine, & Rowland, unpub.), available on the Open Science Framework (at https://osf.io/kau5f/). For this task, children were asked to repeat 24 non-words (6 one-syllable, 6 two-syllable and 6 three-syllable words). At each syllable length, half of the words were wordlike (contained high word average biphone probability) and half of the words were non-wordlike. Words were presented in order of syllable length: 1, 2 then 3 syllables. The task was embedded in a fuzzy felt game in which children attached fuzzy felt animals to a farmyard background. The experimenter produced each non-word a maximum of two times, always in a carrier phrase (e.g. "can you say ..."). The experimenter spoke the non-word and immediately followed this with the presentation of a fuzzy felt picture. When the children attempted the repetition, they were given the picture to attach to the background, regardless of the accuracy of their attempt.

Scores were calculated by totalling the proportion of consonants in each non-word repeated correctly. Consonant errors were coded at the phoneme level. A score of 1 was given for each correct consonant. A score of 0 was given for each omitted or substituted consonant. Added consonants or phonemes and common articulation substitutions were not counted as errors (see Bidgood et al, for details of the coding scheme).

2.2.7 Analysis strategy

Pearson's correlations were performed in R (R Core Team, 2017, R version 3.4.1) using the R Studio (Version 1.0.153) cor.test function, part of the R base package. Regressions were run using the lm function from the lme4 package (Bates, Maechler, Bolker & Walker, 2015). Categorical variables were sum-coded. For sex, girls were coded as 1 and boys were coded as -1. Negative family history of language delay or dyslexia was coded as 1 and positive was coded as -1. Numeric variables were centred for the regression analyses.

We used two cut-offs to classify children as slow talkers at 24 months: having productive vocabulary below the 25th percentile (253 or fewer words produced) and below the 15th percentile (169 or fewer words products) for this sample. The 25th percentile is an inclusive definition and included a number of children who would not be considered late talkers by some criteria (e.g. Rescorla, 1989). However, the use of the 25th percentile is grounded in the literature (Duff et al., 2015) and gave us enough children classified as slow to talk for robust interpretation of analyses. The use of the 15th percentile cut-off is more consistent with previous research on late talkers (Reilly et al., 2010; Hadley & Short, 2005) but this gave us fewer children in our sample of slow talkers.

We constructed receiver operating characteristic (ROC) curves in SPSS Statistics 24 (Hajian-Tilaki, 2013) to determine the sensitivity and specificity of each of our factors at predicting language group at 24 months individually. Next we ran four different discriminant function analyses in SPSS Statistics 24 to determine the power of different combinations of factors to predict language group at 24 months, and another set on language group at 30 months to verify the results on a different outcome measure: 1) all the variables which are successful in predicting unique variance in our regressions; 2) only the non-experimental measures; 3) only the variables from earlier time points, and 4) only non-experimental measures from earlier time points. We chose these four combinations so that we could first identify the predictive power of all the variables combined, then determine if we could exclude costly, time-consuming experimental procedures while retaining the same level of accuracy and finally, to investigate whether accuracy levels are maintained when predicting only from earlier time points. We ran these four analyses twice, first predicting language group membership at 24 months and then predicting language group membership at 30 months.

2.3 Results

Descriptive statistics for each of the numeric variables can be seen in table 2.2 including number, mean and standard deviations. The range is large for all variables, suggesting that there are, indeed, meaningful individual differences in our sample that we can model.

Table 2.2

Descriptive statistics for each numeric variable

Category	Variable	*n* (girls)	*M*	*SD*	Range
Input variables:	Adult word count (*n* per day)	79(41)	18,238.25	7017.04	7,239.00-43,926.50
	Conversational turn count (*n* per day)	79 (41)	806.11	322.58	300.50- 1,804.00
Language variables (speech):	18 month receptive vocabulary	78 (40)	246.60	86.25	66.00-394.00
	18 month productive vocabulary	78 (40)	83.68	69.60	1.00-324.00
	24 month productive vocabulary	79 (41)	371.46	161.67	13.00-679.00
	19 month M3L	72 (37)	1.99	1.15	1.00-6.00
	24 month MLU	77 (40)	1.79	0.52	0.89-3.30
Gesture use at multiple time points:	8 month gestures	72 (36)	10.19	5.69	1.00-28.00
	9 month gestures	64 (31)	14.48	6.35	3.00-34.00
	11 month gestures	67 (34)	21.54	8.13	6.00-47.00
	12 month gestures	72 (37)	27.49	8.28	10.00-55.00
	15 month gestures	77 (41)	40.27	8.64	19.00-62.00
	16 month gestures	75 (39)	46.15	8.44	21.00-63.00
	18 month gestures	78 (40)	50.83	7.40	25.00-63.00
Cognitive variables:	19 month speed of linguistic processing	76 (40)	741.57	139.12	410.25-1,005.33
	25 month non-word repetition	66 (35)	0.65	0.18	0.08-0.96

Note. Adult word count= the number of words spoken by all adults in the child's environment; Conversational turn count= the total number of conversational turns taken, defined as: a child utterance is followed by an adult utterance, or vice versa, within a 5 second window; M3L= mean length of the child's three longest utterances; MLU= mean length of morphemes per utterance; speed of linguistic processing= mean reaction time to look from the distractor picture to target picture; non-word repetition= mean proportion of consonants repeated correctly across all non-words.

2.3.1 Predicting individual differences in productive vocabulary at 24 months

This analysis was run in two steps. In step 1, we ran simple linear regressions (for our two categorical factors) and correlations (for our numerical factors) to test for a relationship with productive vocabulary size at 24 months. Note that we also ran an initial regression to test whether there was an effect of experimenter (there were three experimenters, each of whom was responsible for testing one third of the children) but this was non-significant so is not discussed further. We did not apply corrections for multiple testing, because step 1 was included simply to identify predictors to carry forward to the regression, and we did not want to miss potential predictors by applying too conservative a criterion. Then, in step 2, those factors that were significantly correlated with 24 month productive vocabulary size (all $rs > 0.29$) were entered into multiple regressions to determine whether they remained unique predictors.

The results of the linear regressions on sex and family history of language delay or dyslexia can be seen in Table 2.3. Child sex was a significant predictor of 24 month productive vocabulary, accounting for 21% of the variance in the scores, so this variable was retained at step 2. As girls were coded as 1 and boys were coded as -1, the positive R^2 indicates that being a girl is associated with a larger productive vocabulary. However, family history of language delay or dyslexia was not a significant predictor, so was not retained for step 2.

Table 2. 3

Relationship between predictor variables and 24 month productive vocabulary scores. Statistics given are results of linear regression for categorical predictor variables and correlations for numeric variables

Category	Variable	R^2	*SE*	*t*	*r*	*p*
Demographics	Sex	.21	16.24	4.59		<.001
	Family history	.0002	30.34	.13		.901
Input variables	Conversational turn count				.33	.003
	Adult word count				.13	.26
Language variables (speech):	18 month productive vocabulary				.65	<.001
	18 month receptive vocabulary				.65	<.001
	19 month M3L				.48	<.001
	24 month MLU				.57	<.001
Gesture use at multiple time-points	8 month gestures				.13	.28
	9 month gestures				.17	.19
	11 month gestures				.32	.009
	12 month gestures				.29	.01
	15 month gestures				.40	<.001
	16 month gestures				.40	<.001
	18 month gestures				.45	<.001
Cognitive variables	19 month speed of linguistic processing				-.46	<.001
	25 month NWR				.58	<.001

The list of our numeric predictor variables and the results of the correlations between each of these variables and our outcome measure (24 month productive vocabulary scores) can be seen in Table 2.3. Only one of our input variables correlated with 24 month productive vocabulary size – conversational turn-count, so this was retained. All four spoken language variables (18 month productive vocabulary size, 18 month receptive vocabulary size, 19 and 24 month MLU) correlated significantly with 24 month productive vocabulary size, with large effect sizes.

Early gesture use (at 8 and 9 months) was not significantly correlated with 24 month productive vocabulary, but gesture use from 11 months onwards was. However, gesture scores across time were collinear at most data points between 11 and 18 months of age (see table 2.4), making it difficult to include them all in a regression analysis. Thus, we chose to retain only gesture scores from 11 and 18 months of age, which were only moderately correlated (r =.55) with VIF values under 10 (11 month gestures VIF=1.41; 18 month gestures VIF=1.41; Alin, 2010). Finally, both of our cognitive variables – 19 month speed of processing and 25 months non-word repetition performance – were significantly correlated with 24 month productive vocabulary size, so both were retained. Note that the effect size for speed of processing is negative (r = -.46), indicating that, as predicted, faster processers (smaller reaction time) had larger productive vocabulary scores.

Table 2.4

Collinearity results for gesture scores at 11, 12, 15, 16 and 18 months

Age	11 months	12 months	15 months	16 months	18 months
11 months	-				
12 months	.82***	-			
15 months	.72***	.77***	-		
16 months	.69***	.76***	.85***	-	
18 months	.55***	.63***	.72***	.84***	-

We also ran correlation analyses between each of the predictor variables that correlated significantly with 24 month productive vocabulary, to check if any of the variables were highly collinear. None of the nine variables correlated highly; all were below *r*=0.80 and therefore were kept for the regression analyses. See table 2.5 for the results of these correlations.

Table 2.5

Collinearity results for all numeric predictor variables that correlated with 24 month productive vocabulary scores

Variable	Conversational turn count	18 month productive vocabulary	18 month receptive vocabulary	19 month M3L	24 month MLU	11 month gestures	18 month gestures	19 month speed of linguistic processing	25 month NWR
Conversational turn count	-								
18 month productive vocabulary	0.43	-							
18 month receptive vocabulary	0.17	0.58	-						
19 month M3L	0.45	0.72	0.47	-					
24 month MLU	0.41	0.58	0.43	0.55	-				
11 month gestures	-0.10	0.45	0.48	0.43	0.24	-			
18 month gestures	0.04	0.51	0.52	0.42	0.35	0.55	-		
19 month speed of linguistic processing	-0.29	-0.36	-0.35	-0.35	-0.42	-0.04	-0.24	-	
25 month NWR	0.41	0.46	0.41	0.36	0.51	0.20	0.29	-0.45	-

In step 2, we ran regression analyses with all the variables that we retained from step 1. We were particularly interested in whether any of our predictor variables remained significant after accounting for the variance explained by sex and previous productive vocabulary size (at an earlier time point, in our case 18 months), which have already been identified in the previous literature as robust predictors of vocabulary size. Thus, we first entered sex and productive vocabulary size at 18 months into our base model, and then, for each model, we individually entered the additional predictor variable that we retained from step 1 above: conversational turn count, 18 month receptive vocabulary, 19 and 24 month MLU, 11 and 18 month gestures, speed of linguistic processing, and NWR.

The results of the base model, including both sex and 18 month productive vocabulary can be seen in table 2.6, and the results for the additional eight regressions, each of which contained the baseline variables plus one additional variable, are in table 2.7. Again, we chose not to correct for multiple comparisons in order to retain potentially influential variables for our sensitivity and specificity analysis below, but note that applying a conservative Bonferroni correction would reduce the critical p value to 0.006.

Table 2.6

Base multiple regression model with Sex and 18 month productive vocabulary as predictors and 24M productive vocabulary size as the outcome measure.

Base Model	R^2	Adj. R^2	*F*	*df*	*p*
Sex + 18 month productive vocabulary	.49	.48	36.62	2,75	<.001
		B	*SE*	*t*	*p*
Sex		47.52	14.03	3.39	.001
18 month productive vocabulary		1.25	0.20	6.15	<.001

Note. Adj. =Adjusted

The base model alone explains a significant proportion of the variance in 24M productive vocabulary scores (49%) but additional significant variance was explained by both spoken language measures (18 month receptive vocabulary explained an additional 9% variance, and MLU at 24 months an additional 5% variance) and both cognitive predictors (speed of processing predicted an additional

6% and NWR performance an additional 5% variance). However, M3L at 19 months, both gesture scores, and conversational turn count were not unique significant predictors of variance in productive vocabulary scores at 24 months.

Table 2.7

Results of eight separate multiple regression models. Each model contains the baseline predictor variables plus one additional predictor variable, specified below

Model	Variables	R^2	Adj. R^2	*F*	*df*	*p*
Base model + conversational turn count		.50	.48	24.70	3,74	<.001
			B	*SE*	*t*	*p*
	Sex		49.79	14.24	3.50	<.001
	18 month productive vocabulary		1.14	0.23	4.97	<.001
	Conversational turn count		0.04	0.05	0.96	.34
Base model+18 month receptive vocabulary		R^2	Adj. R^2	*F*	*df*	*p*
		.58	.56	33.84	3,74	<.001
			B	*SE*	*t*	*p*
	Sex		37.33	13.17	2.84	.006
	18 month productive vocabulary		0.81	0.22	3.71	<.001
	18 month receptive vocabulary		0.68	0.18	3.85	<.001
Base model+ 19 month M3L		R^2	Adj. R^2	*F*	*df*	*p*
		.52	.50	23.99	3,67	<.001
			B	*SE*	*t*	*p*
	Sex		57.28	14.62	3.92	<.001
	18 month productive vocabulary		1.11	0.29	3.83	<.001
	19 month M3L		2.95	17.21	0.17	.86
Base model+ 24 month MLU		R^2	Adj. R^2	*F*	*df*	*p*
		.54	.52	27.72	3,72	<.001
			B	*SE*	*t*	*p*
	Sex		45.73	13.87	3.30	.002
	18 month productive vocabulary		0.90	0.24	3.83	<.001

	24 month MLU		81.02	30.59	2.65	.01[!]
Base model+ 11month gestures		R^2	Adj. R^2	*F*	*df*	*p*
		0.52	.50	22.74	3,62	<.001
			B	*SE*	*t*	*p*
	Sex		46.96	15.47	3.04	.004
	18 month productive vocabulary		1.37	0.25	5.45	<.001
	11 month gestures		1.07	2.02	0.53	.60
Base model+ 18 month gestures		R^2	Adj. R^2	*F*	*df*	*p*
		.51	.49	25.18	3,74	<.001
			B	*SE*	*t*	*p*
	Sex		42.90	14.42	2.98	.004
	18 month productive vocabulary		1.14	0.22	5.23	<.001
	18 month gestures		2.66	2.06	1.29	.20
Base model+ speed of linguistic processing		R^2	Adj. R^2	*F*	*df*	*p*
		.55	.53	28.83	3,71	<.001
			B	*SE*	*t*	*p*
	Sex		46.94	13.61	3.45	.001
	18 month productive vocabulary		1.06	0.20	5.16	<.001
	Speed of linguistic processing		-0.22	0.10	-2.16	.03[!]
Base model+ non-word repetition		R^2	Adj. R^2	*F*	*df*	*p*
		.54	.52	23.86	3,61	<.001
			B	*SE*	*t*	*p*
	Sex		31.96	14.21	2.25	0.03[!]
	18 month productive vocabulary		0.91	0.21	4.36	<.001
	Non-word repetition		243.01	84.09	2.89	.005

Note: Adj. =Adjusted
[!] Does not reach critical p of < .006 if Bonferroni correction is applied.

We ran one final regression including all significant predictors from the last round. The full model explained 63% of the variance in productive vocabulary scores, though only 18 month receptive vocabulary remained as a significant unique predictor (see Table 2.8)[1].

Table 2.8

Full regression model predicting variance in 24 month productive vocabulary scores

Model	Variables	R^2	Adj. R^2	*F*	*df*	*p*
Full model		.63	.59	15.45	6,55	<.001
			B	*SE*	*t*	*p*
	Sex		23.24	13.46	1.73	.09
	18 month productive vocabulary		0.41	0.24	1.72	.09
	18 month receptive vocabulary		0.60	0.19	3.20	.002
	24 month MLU		13.49	29.41	0.46	.65
	Speed of processing		-0.12	0.10	-1.12	.27
	Non-word repetition		163.67	84.99	1.93	.06

Adj. =Adjusted

2.3.2 Classifying children who are slow to learn to talk at 24 months using ROC curves

The aim of this analysis was to determine how accurately each of our predictor variables was able to classify children in our slow-to-talk sample. The 15th percentile cut-off yielded 12 children and the 25th percentile cut-off yielded 20 children in our slow to talk sample. Our predictor variables were those that were significantly related to productive vocabulary in our step 1 analysis above: 18 month productive vocabulary, 18 month receptive vocabulary, MLU at 24 months, speed of linguistic processing and NWR performance. As a receiver operating characteristic (ROC) curve analysis assesses the accuracy of a continuous measure for predicting a

[1] In the interests of full reporting, note that we also ran growth curve analyses (Mirman, 2014) to predict growth in productive vocabulary between 24, 27 and 30 months, since these were in our original analysis plan. However, ceiling effects in our outcome data between 25 and 30 months meant we could not sensibly interpret the results of the linear term, so these results are not reported here. For this analysis, see supplementary materials at https://osf.io/2xgtv/.

binary outcome, sex could not be included in this analysis but is included in the discriminant function analyses below.

We constructed ROC curves to determine the diagnostic classification accuracy of each of our predictor variables individually. The ROC curve is created by plotting the true positive rate (or sensitivity) against the false positive rate (or 1-specificity) at various threshold settings. This yields an area under the curve (AUC) measure that can be used to judge the overall accuracy of the predictor at classifying the participants into the two groups. AUC values range from 0.5 to 1, and higher AUC values indicate better classification accuracy. The analysis also yields sensitivity and specificity measures. Sensitivity measures the ability of the predictor to correctly identify the children in our slow to talk group (the true positive rate). Specificity is the ability of the predictor to correctly identify children who are not in our slow to talk group (true negative rate). While there are no gold standard sensitivity and specificity rates for classifying late talkers, sensitivity and specificity rates between 70-80% are considered good for diagnostic assessments for autistic spectrum disorder (Bright Futures Steering Committee & Medical Home Initiatives for Children with Special Needs Project Advisory Committee, 2006). Thus we consider accuracy rates above 70% to be good in the present analysis.

The results from the ROC curve analyses at 24 months can been seen in tables 2.9 and 2.10. These tables also detail the best cut-off scores in terms of maximising both sensitivity and specificity. Using the 25th percentile cut-off, the best cut-off scores were: for productive vocabulary at 18 months, 45.5 words or less (sensitivity 83%, specificity 81%), for receptive vocabulary at 18 months: 206 words or less (sensitivity 83%, specificity 81%); for MLU at 24 months: MLU under 1.53 or less (sensitivity 90%, specificity 83%); for NWR at 25 months: 59% or fewer consonants repeated accurately (sensitivity 85%, specificity 81%).

The speed of processing predictor was less successful. The AUC score was below 0.80, and even the best cut-off score (a reaction time of 744.79ms or slower) only identified slow talkers at 24 months with a sensitivity of 82% and a specificity of 57%.

Using the 15th percentile cut-off, the best cut-off scores were: for productive vocabulary at 18 months, 38 words or less (sensitivity 91%, specificity 82%), for receptive vocabulary at 18 months: 196 words or less (sensitivity 91%, specificity 79%); for MLU at 24 months: MLU under 1.47 or less (sensitivity 92%, specificity

81%); for speed of processing: a reaction time of 820.43ms or slower (sensitivity 80%, specificity 81%); for NWR at 25 months: 59% or fewer consonants repeated accurately (sensitivity 100%, specificity 76%). The results of these analyses can be seen in table 2.10. These results yielded similar sensitivity and specificity values to the analyses using the 25th percentile cut-off, thus, in the interests of brevity, for the remainder of the analyses we use only the 25th percentile cut-off.

Table 2.9

ROC Analyses Results: Diagnostic classification accuracy for each predictor variable correctly distinguishing between children with and without language delay using the 25th percentile cut-off

Variable	Best cut-off	AUC	SE	*p*	Sensitivity	Specificity
18 month productive vocabulary	≤45.5 words	0.90	0.04	<.001	0.83	0.81
18 month receptive vocabulary	≤206 words	0.81	0.06	<.001	0.83	0.81
24 month MLU	≤1.53 morphemes	0.89	0.04	<.001	0.90	0.83
Speed of Processing	≥744.79ms	0.73	0.07	.005	0.82	0.57
Non-word repetition	≤59% accuracy	0.87	0.06	<.001	0.85	0.81

Table 2.10

ROC Analyses Results: Diagnostic classification accuracy for each predictor variable correctly distinguishing between children with and without language delay using the 15th percentile cut-off

Variable	Best cut-off	AUC	SE	*p*	Sensitivity	Specificity
18 month productive vocabulary	≤38 words	0.89	0.06	<.001	0.91	0.82
18 month receptive vocabulary	≤196 words	0.83	0.08	<.001	0.91	0.79
24 month MLU	≤1.47 morphemes	0.89	0.05	<.001	0.92	0.81
Speed of Processing	≥820.43ms	0.79	0.07	0.005	0.80	0.81
Non-word repetition	≤59% accuracy	0.94	0.04	<.001	1.00	0.76

2.3.3 Classifying children who are slow to learn to talk at 24 and 30 months using discriminant function analysis

We ran eight discriminant analyses, with four different combinations of variables, to determine whether combinations of our predictor variables give us better classification accuracy at predicting language group at 24 months than each predictor individually. We then repeated these analyses at 30 months, to test the pattern of results against a different, but related, outcome measure. We recoded our predictor variables into categorical variables (1, 2) based on the best cut-off identified by the ROC analysis above. For example, the best cut-off for the 18 month productive vocabulary predictor was *producing 45.5 or fewer words*. Thus, we created a new variable in which children producing 45.5 words or fewer were recoded as 1, and children producing more than 45.5 words were recoded as 2. We then used the 25th percentile cut-off for productive vocabulary at 30 months to create language groups at 30 months. This cut-off was producing fewer than 527 words and gave us 18 children in our sample of slow talkers. For analyses using language groups at 30 months, we still used cut-off scores from the ROC analyses created using the language groups at 24 months to identify if these cut-off scores remained strong predictors over time.

The first discriminant function analysis included all of the factors which were included in the sensitivity and specificity analyses above: 18 month productive vocabulary, 18 month receptive vocabulary, MLU at 24 months, NWR performance and speed of linguistic processing, as well as child sex. This model correctly classified children into their language groups at 24 months using the combination of all the factors with an accuracy rate of 91.50% with a sensitivity of 81.80% and specificity of 93.80%. The results of this analysis can be seen in table 2.11.

This model also correctly classified children into their language groups at 30 months, with an overall accuracy of 86.00%, and specificity of 93.50%. However, sensitivity was only 54.50%. Thus, while the accuracy of this model was good, it was not successful in correctly classifying children whose productive vocabulary was below the 25th percentile at 30 months.[2]

[2] Note that we also ran a discriminant function analysis including all these variables as well as family history of language delay or dyslexia since this was in our original analysis plan. We chose to run this analysis on the basis that family history might be a good classifier of slow to talk children

The second discriminant function analysis included only our non-experimental measures: sex, 18 month productive vocabulary, 18 month receptive vocabulary and 24 month MLU. Experimental tasks are time-consuming and expensive to run, so it is important to know whether they add any explanatory power. This model correctly classified children into their language groups at 24 months at a very high rate of 91.90%, with a sensitivity of 83.30% and a specificity of 94.60%. Comparison of accuracy measures for the full model and the non-experimental model revealed only a 0.4% difference in accuracy, suggesting that, while the exclusion of experimental measures slightly increases the model's accuracy, it provides very little additional explanatory power.

This model also correctly classified children into their language groups at 30 months with an accuracy of 80.00%, specificity 84.90%. However, sensitivity was poor (64.70%). In sum, this model was 6% less accurate that the full model at 30 months, which means that the inclusion of experimental factors increased the model's ability to classify children who are, and are not, slow to learn to talk at 30 months.

The third discriminant function analysis only included variables from earlier time points, before 24 months, to determine whether we can use earlier measures to identify later group membership: sex, 18 month productive vocabulary, 18 month receptive vocabulary and 19 month speed of processing. This model correctly distinguished between language groups at 24 months with an accuracy of 89.20%, sensitivity of 88.20% and specificity of 89.50%. This model was only slightly less accurate than the first model that had all the variables included (91.50% vs 89.20%).

This model also correctly distinguished between language groups at 30 months with an accuracy of 84.50%. The time, both the sensitivity and specificity of this model were also good, 81.30% and 85.50% respectively. Overall, this model was only 1.50% less accurate than the full model (model 1 above, 86.00% vs 84.50%), and the sensitivity was much better, meaning this model was better at correctly classifying children whose productive vocabulary was below the 25th percentile at 30 months.

---

even if it is not a good predictor of individual differences. The inclusion of this variable had no effect on the result.

Finally, the fourth discriminant function analysis included only non-experimental variables from earlier time points. This model correctly identified language group membership at 24 months using these factors with an accuracy of 88.50%. However, while the specificity of this model was good (93.20%) the sensitivity was not as good (73.70%), meaning it was less successful in correctly classifying slow-to-talk children when compared to model 3, which also included experimental measures (speed of linguistic processing; sensitivity = 88.20%). Therefore, for classifying groups based on productive vocabulary at 24 month using measures only from earlier time points, speed of processing is a useful additional measure to have, over and above productive and receptive vocabulary.

This model also correctly distinguished between language groups at 30 months with an accuracy of 79.50%. However, while the specificity of this model was good (85.50%), the sensitivity was poor (61.10%). This means that this model did not do well in terms of correctly classifying children whose productive vocabulary was below the 25^{th} percentile at 30 months. Therefore, the inclusion of speed of linguistic processing (model 3 above) again improves our ability to classify children into language groups at 30 months using data from before 24 months.

The standardised canonical correlations for each model can be seen in table 2.12. These values provide the contribution of each variable in classifying children into their groups.

Table 2.11

Results of the discriminate function analyses for various combinations of variables for predicting language group membership at 24 and 30 months

Variable	*Month*	*r*	χ^2	*n*	*df*	*p*	Accuracy	Sensitivity	Specificity
All variables (sex, 18 month productive & receptive vocabulary, 24 month MLU, speed of processing, non-word repetition)	24	.72	38.74	59	6	<.001	91.50%	81.80%	93.80%
	30	.55	18.51	57	6	.005	86.00%	54.50%	93.50%
No experimental data (sex, 18 month productive & receptive vocabulary, 24 month MLU)	24	.77	63.31	74	4	<.001	91.90%	83.30%	94.60%
	30	.58	27.06	70	4	<.001	80.00%	64.70%	84.90%
Only earlier data (Sex, 18 month productive & receptive vocabulary, speed of processing)	24	.71	48.12	74	4	<.001	89.20%	88.20%	89.50%
	30	.62	32.09	71	4	<.001	84.50%	81.30%	85.50%
Earlier Vocabulary (Sex, 18 month productive & receptive vocabulary)	24	.68	46.73	78	3	<.001	88.50%	73.70%	93.20%
	30	.58	28.68	73	3	<.001	79.50%	61.10%	85.50%

Table 2.12

Standardised canonical coefficients for all variables in the four different discriminant function analyses predicting language group at 24 months

Model	Variable	*r* (24 months)	*r* (30 months)
All variables	Sex	.03	.002
	18 month productive vocabulary	.48	.18
	18 month receptive vocabulary	.36	.52
	24 month MLU	.55	.05
	Speed of linguistic processing	.32	.46
	Non-word repetition	.09	.30
Non-experimental	Sex	-.07	-.25
	18 month productive vocabulary	.41	.11
	18 month receptive vocabulary	.50	.74
	24 month MLU	.67	.34
Only Earlier	Sex	-.11	-.12
	18 month productive vocabulary	.72	.32
	18 month receptive vocabulary	.42	.69
	Speed of linguistic processing	.38	.37
Earlier non-experimental	Sex	-.14	-.29
	18 month productive vocabulary	.63	.24
	18 month receptive vocabulary	.61	.79

2.4 Discussion

The goal of the present study was to investigate whether 10 factors, all implicated in explaining individual differences in vocabulary size in the first two years of life, could also be used to discriminate between children who were, and were not, slow to learn to talk, in a sample of 79 children aged 24 and 30 months of age. We first assessed whether our measures predicted individual differences in 24 month productive vocabulary development using correlational and simple linear regression analyses. In line with our predictions, 18 month productive vocabulary and sex were significant predictors of 24 month productive vocabulary and were therefore entered as baseline predictors for subsequent analyses. In addition, one input measure (conversational turn count), three language measures (18 month receptive vocabulary, MLU at both 19 and 24 months, and gesture scores at 11, 12, 15, 16 and 18 months), and both cognitive measures (19 month speed of linguistic processing, 25 month NWR) correlated with 24 month productive vocabulary development. However, only two language measures (18 month receptive vocabulary, 24M MLU) and the two cognitive measures (speed of linguistic processing, NWR) remained unique predictors when controlling for 18 month productive vocabulary and sex.

Three of our null findings - family history, adult word count and gesture scores at 8 and 9 months - can be explained relatively easily. First, although a family history of speech or language impairment has been shown previously to be a robust predictor or individual differences in vocabulary development (Reilly et al., 2007; Reilly et al., 2009; Zubrick et al, 2007), our sample probably did not contain enough variance to capture an effect; only nine children, out of 79, had a family history of speech or language impairment. Second, adult word count, as measured by LENA, includes both overheard speech and child directed speech, and an increasing body of evidence shows that while children can learn from overhearing (Akhtar, Jipson, & Callanan, 2001), it may be child directed speech that matters for vocabulary development, and thus predicts vocabulary size (Weisleder & Fernald, 2013). In addition, it may not be quantity but quality (lexical diversity) of child directed speech that matters for vocabulary development, which we did not measure here (Cartmill et al., 2013; Jones & Rowland, 2017; Vigil, Hodges, & Klee, 2005). Third, the finding that gesture scores at 8 and 9 months did not correlate with productive vocabulary at 24 months is most likely due to a floor effect; children are not producing enough gestures at 8 and 9 months for us to find a relationship with productive vocabulary.

More interesting is the failure to find an effect of conversational turn counts once we added productive vocabulary at 18 months and child sex into the regression model, given the research supporting its role in vocabulary development (Romeo et al., 2018; Zimmerman et al., 2009). However, the result becomes less surprising if you characterise conversational turn taking not as a pure measure of the linguistic environment (e.g. a measure of how much the parent allows the child to take a turn in the conversation) but as a measure, at least in part, determined by the child (e.g. a measure of how talkative a child is). Children with larger vocabularies are likely to talk more, and are thus likely to take more turns in the conversation. This means that relationships between conversational turn taking between 18-21 months and vocabulary scores at 24 months is likely to be at least partly mediated by vocabulary scores at 18 months, as is the case here. Note that conversational turn counts and child talkativeness are not exactly the same thing. For example, Romeo et al. (2018) found a relationship between turn taking and vocabulary size when controlling for talkativeness. However, conversational turn taking behaviour may, in large, be influenced by the child's vocabulary size, rather than a pure measure of the quality of the linguistic environment.

Similarly surprising is the fact that gesture use at earlier time points did not uniquely predict subsequent vocabulary size. Although the correlations between gesture use at 11, 12, 15, 16 and 18 months and productive vocabulary at 24 months were significant, the effects of gesture use disappeared in the regression when controlling for productive vocabulary at 18 months and sex. However, the result becomes less surprising when we consider the fact that previous studies that have found robust relationships between early gesture use and later vocabulary (Rowe et al., 2008; Rowe & Goldin-Meadow, 2009) do not, in the main, control for vocabulary at prior data points. While gesture use may be a predictor of later vocabulary development, it is likely highly correlated with concurrent vocabulary, and thus does not predict unique variance when controlling for concurrent vocabulary or sex. Overall though, we replicated the effect of four predictors from the literature on individual differences in vocabulary size at 24 months; the effect of earlier receptive vocabulary at 18 months, of concurrent language complexity (MLU at 24 months), of earlier speed of linguistic processing ability (at 19 months) and of concurrent non-word repetition ability (at 25 months).

Our second set of analyses established the discriminatory ability of our six factors in classifying children as slow to talk at 24 and 30 months of age. When distinguishing between language groups at 24 months, using both 25th and 15th percentile cut-off points, most of the variables that were successful at predicting individual differences (18 month productive and receptive vocabulary, MLU at 24 months, speed of processing and NWR) were successful in identifying children in the slow-to-talk group, with acceptable levels of sensitivity. However, speed of processing was less successful in identifying children who were not slow to talk (specificity was low).

The findings from our discriminant function analyses at 24 months show that cut-off scores from all of these variables can be used in combination to successfully distinguish between children whose productive vocabulary was above and below the 25th percentile at 24 months. In addition, the model using non-experimental measures alone (sex, 18 month productive and receptive vocabulary and MLU at 24 months) was equally accurate in distinguishing between these children. Similarly, using earlier measures only (18 month productive and receptive vocabulary and speed of processing) was successful in separating the two groups. However, when using only earlier non-experimental data, we were less successful in identifying children who were slow to talk. To successfully classify children based on their productive vocabulary at 24 months, using only measures from earlier time points, a measure of speed of processing improves our prediction.

The full model was less successful in identifying children who were slow talkers at 30 months, as were the models using only language measures (models 2 and 4). However, model 3, that included only 18 month vocabulary and 19 month speed of processing measures was successful in identifying both children who were and were not slow to talk, with good sensitivity and specificity. Thus, adding 19 month speed of processing to early vocabulary (productive and receptive) measures adds additional explanatory power when predicting long term (to 30 months) growth.

Overall, considering both individual differences and classification analyses together, the most consistently robust predictor of individual difference in vocabulary development and of language group at 24 and 30 months, was the child's own communicative ability, both at an earlier time point (18 month vocabulary predicting language group at 24 and 30 months) and concurrently (24 month MLU predicting language group at 24 months). This is consistent with previous research showing that earlier vocabulary measures are consistently our most robust predictors, both for predicting individual differences (Henrichs et al., 2011; Westerlund et al., 2006) and late talking status (Klee et al., 1998; Reilly et al., 2007).

Language use at an earlier time point is likely to influence later vocabulary for a number of reasons. First, even if the rate of growth is static, children who start out with more words at 18 months will end up with more words at 24 months and 30 months. Second, rate of growth is also likely to be faster, since all the factors that led to children being fast at the earlier time point are presumably still operating. Third, and most interestingly, is the idea that the amount a child knows at one time point might actually facilitate faster subsequent vocabulary growth, an idea suggested by a number of researchers. For example, Borovsky, Ellis, Evans and Elman, (2016) have suggested that the more words a child knows the more semantic concepts they have, the easier they can recognise and thus learn new words. Therefore, children who know more words are quicker to learn new words because they have more semantic concepts to refer to compared to children who know fewer words and thus have fewer semantic concepts. Alternatively, Jones & Rowland (2017) suggest that children who know a lot of words have a bigger store of sublexical chunks in long term memory that they can call upon to process and learn new material. This means that children with a bigger store of sublexical chunks should therefore learn subsequent words faster than children who know fewer words and thus have a smaller store of sublexical chunks to refer to when they encounter new words.

This idea that the size of a child's lexicon affects the speed with which they build subsequent vocabulary faster, gains additional support from the fact that, in our data, receptive vocabulary at 18 months was such a robust predictor even when we took into account productive vocabulary at 18 months. It added substantial additional variance even in the full regression model which included productive vocabulary at 18 months, and was consistently one of the most reliable variables for classifying children into our slow to talk group. This finding - that the number of words that a child knows influences later productive vocabulary growth over and above the number of words they can produce at the same time point - suggests that lexicon size may play an important role in later vocabulary acquisition.

The role of our cognitive predictors - speed of processing and non-word repetition performance - is less clear cut. Both of these have been suggested in the literature to be strong predictors of later vocabulary skills (Fernald et al., 2006; Fernald & Marchman, 2012; Killing & Bishop, 2008 for speed of processing; Duff et al., 2015; Hoff et al., 2008; Thal et al., 2005 for NWR), which makes them excellent candidates for identifying children who are slow to talk. NWR performance, in particular, is said to be a strong predictor for later developmental language disorder (Bishop, North & Donlan, 1996; Gathercole & Baddeley, 1990b).

While we replicated the findings that speed of processing predicts later vocabulary size, it consistently had less predictive power than our language measures. In addition, speed of processing did not predict unique variance in our full regression model, which also included all language measures, sex and NWR performance, and it was one of the least accurate predictors at classifying our slow to talk group at 24 months. That said, it substantially improved the accuracy of the classifier model that was given only 18-19 month old data to predict language group at both 24 and 30 months. Therefore, our data suggest that processing speed adds additional explanatory power to the ability to predict, from 19 months of age, which children are likely to have productive vocabulary scores below the 25th percentile. However, why this is still remains to be explained. Some speech perception models propose that increases in vocabulary size allow for more efficient, perceptual routines for word recognition, which contributes to faster lexical processing speeds (Curtin & Werker, 2007; Strange, 2011). There is also the possibility that children with larger vocabularies have more robust lexical knowledge, and more highly-developed lexical representations. This would lead to more rapid, efficient processing of known words relative to children with smaller vocabularies, even if words are familiar to both groups of children, which would free up cognitive resources for the more efficient encoding of new words. Both these explanations are consistent with our findings, so further work is necessary to disentangle cause and effect in speed of processing and vocabulary size studies.

We also replicated findings that show NWR predicts variance in productive vocabulary size and expanded on this by showing that NWR uniquely predicts vocabulary when controlling for earlier productive vocabulary and child sex. However, it did not predict unique variance in our full regression model and did not seem to substantially add to the accuracy of our classification models. In addition, we can see from the discriminate function analysis table 2.11, the model that includes NWR has fewer participants. More children who did not complete the NWR task had lower vocabulary scores. Thus, NWR may not be practical for identifying children whose vocabulary scores are delayed due to missing data.

Jones, Gobet and Pine (2007) propose that NWR examines how phonological working memory and long-term memory interact; performance on NWR tasks improves with age not because of increases in phonological working memory but because of increases in the amount of information stored in the long-term memory. If NWR is largely a reflection of vocabulary size (information stored in the long-term memory), then it may not add much explanatory power to a model that already contains vocabulary measure. We need to know more about what it measures, and about the relationship between non-word repetition performance and vocabulary size, before we draw causal conclusions about its predictive power to identify children who are slow to talk.

2.5 Conclusion

We found that 18-19 month receptive vocabulary and speed of linguistic processing scores, and 24-25 month MLU and NWR scores were robust predictors of individual differences in 24 month productive vocabulary scores. These measures successfully predicted variance in productive vocabulary scores when earlier productive vocabulary and sex were controlled for. They also successfully distinguished between children who were, and were not, slow to learn to talk. We conclude that these measures may be useful tools to monitor children's language development from an early age.

3 Chapter 3: Individual differences in vocabulary acquisition in late talking and typically developing children.

Fit within the thesis

Chapter 2 demonstrated the role of multiple factors in predicting unique variance in productive vocabulary development at 24 months in a sample of typically developing children. Chapter 3 examines the role of these factors in predicting unique variance in productive vocabulary at 25 months in an expanded sample of late talking and typically developing children. Following this, the discriminatory ability of these factors to distinguish between late talking and typically developing children was examined. As outlined in chapter 1, much of the research examining late talking focusses either on language outcomes of late talking children or on distal factors contributing to an increased likelihood of late talking. This chapter examines the role of proximal factors in identifying late talking children to begin to establish the underlying causes of late talking.

The study reported in this chapter found that multiple factors (earlier receptive vocabulary, mean length of children's three longest utterances, gestures and non-word repetition) all explain unique variance in productive vocabulary at 25 months after controlling for child sex and earlier productive vocabulary. In addition, individually these factors can identify late talking and typically developing children. However, additively, they are not successful at discriminating between these two groups of children.

This study was designed by Lana Jago and Caroline Rowland. Lana Jago recruited and collected data from the primary data set of 36 late talking children. This was completed between 2017-2019. All of these children came to The University of Liverpool to complete experimental and standardised language-based assessments. The additional comparison data from typically developing children is the same as that reported in chapter 2, and was collected by Michelle Peter, Amy Bidgood and Samantha Durrant as part of the Language 0-5 Project. Lana Jago coded the primary data, ran the data analyses and wrote the first draft of the paper. Caroline Rowland gave comments and revisions on the paper. This chapter is in preparation for submission to a peer reviewed journal (Jago, Peter, Bidgood, Durrant, Pine & Rowland, in prep.).

Abstract

This study investigated why some children are late talkers and others are not, using a sample of 113 children tested from 15-25 months. These children were part of two samples. The main data was from a sample of 36 children recruited specifically to have lower productive vocabulary scores. The comparison sample of 77 typically developing children was from the Language 0-5 Project. First, we established the best predictors of individual differences in productive vocabulary at 25 months. Then we investigated if the best predictors of individual differences in productive vocabulary could also discriminate between late talking and typically developing children. Regression analyses revealed that earlier receptive vocabulary and gestures from 15-18 months, as well as mean length of the three longest utterances (M3L) and non-word repetition (NWR) at 25 months all predicted unique variance in productive vocabulary after controlling for child sex and earlier productive vocabulary. Receiver operation characteristic curve analyses revealed that the measures can also be used to identify late talking children but discriminant function analyses revealed these measures were less successful, when combined, at correctly classifying late talking children with poor levels of sensitivity. Overall, NWR was the best predictor for identifying late talking children and for discriminating between late talking and typically developing children.

3.1 Introduction

Up to 19% of 2-year-old children are categorised as late talkers (Carson & Gavin, 1998; Collisson et al., 2016; Rescorla, 1989; Rescorla & Achenbach, 2002; Zubrick et al., 2007). While there are many reasons why children express a delay in productive vocabulary development, late talking children do so in the absence of any other cognitive, visual or hearing impairments. Late talkers are identified at approximately 2 years old. There are two main criteria used to identify late talkers: first, late talkers are toddlers who are producing fewer than 50 words; second, late talkers are toddlers who are not combining any words together (Klee et al., 1998; Rescorla, 1989).

Currently, little is known about why some children are late talkers. However, a delay in vocabulary acquisition is associated with an increased risk of later language impairment. Up to half of late talking children remain delayed when followed up in early childhood (Rescorla & Schwartz, 1990) and a substantial proportion continue to perform poorly on measures of vocabulary and grammar in later childhood (Bishop & Edmundson, 1987; Girolametto, Wiigs, Smyth, Weitzman, & Pearce, 2001). For example, Rescorla and Schwarts (1990) followed up a sample of 25 boys from 24 months to 4 years. All 25 boys were diagnosed as having a delay in productive vocabulary at 24 months. At follow-up, half of these children still had delays with productive vocabulary development. However, as it stands, we are currently unable to predict which children's language will catch up and which children will remain delayed (Bishop et al., 2012; Rescorla, 2011).

One of the reasons why we know very little about why some children are late talkers is that most research on late talking has focused on the language outcomes of late talkers, not the predictors of late talking itself (Dale, Price, Bishop, & Plomin, 2003; Rescorla, 2005, 2009). For example, Rescorla (2005) looked at the language outcomes of late talking children at both 13 years and 17 years (Rescorla 2009). While these children's performance on measures of language related skills was below that of their typically developing peers at 13 and 17 years, when these children were identified as late talkers at 2 years, only their vocabulary was assessed. Therefore, it is not possible to know how these children may have differed on factors that are underlying language development at 2 years. Having a better understanding of the causes of late talking is an important step, both for answering the question of why children differ in their early language development in itself, but also in terms of delivering better predictions about which late talkers will later catch up with their peers, and which children will have persisting language impairments.

There are many reasons why children may exhibit a delay in their early productive vocabulary development. For example, it is possible that children with smaller phonological working memories have less opportunity to transfer representations of novel words from short term memory to long term memory and thus, are slower to learn vocabulary (Thal, Miller, Carlson, & Vega, 2005). Alternatively, late talking children may be slower to process familiar words, and thus, have less opportunity to learn new words (Fernald & Marchman, 2012). If we can establish whether late talking children perform differently on such measures, we will begin to understand the underlying causes of late talking. There are also large individual differences in the rate of vocabulary development; children vary in the rate they begin to produce their first words and in the rate they begin to combine words together (Fenson et al., 1994; Kidd, Donelly, & Christiansen, 2018). Many of the factors which explain individual differences in vocabulary development (such as demographic factors: child sex, family history of speech or language impairment; environmental factors: input and conversational turns; language proficiency factors: earlier vocabulary and mean length of utterances; gesture skills; and cognitive factors: speed of linguistic processing and non-word repetition) could also explain late talking. It is likely that measures which can explain why children acquire language at varying rates can also explain why children have a delay in vocabulary acquisition.

However, it is also possible that these measures may only be successful in predicting individual variation but do not distinguish between late talking and typically developing children. While these factors have been shown to play a role in language delay (Fernald & Marchman, 2012; Reilly et al., 2010) there may be too much overlap between scores obtained by typically developing and late talking children for these factors to discriminate well between children. Thus, it is important to investigate whether these factors can also identify late talking children; if we can identify whether or not causes of individual variation in language acquisition can distinguish between typically developing and late talking children, we can use these factors to begin to identify children who may experience a delay in vocabulary acquisition from an early age.

For this study, we used two data sets. First, we used a longitudinal design to identify a sample of potential late talkers by measuring children's vocabulary acquisition at 15-18 months of age. We then tested them again at 25 months of age. At both time points, we collected data on the factors that had been identified in a previous study (Jago et al, in submission, see chapter 2), as predictors of individual differences in language acquisition, and thus are promising predictors of late talking status. To provide a comparison sample of typically developing children, we used the data from the Language 0-5 Project, in which data on multiple factors associated with language development was collected at multiple time points from 6-25 months (the same data used in Jago et al, in submission, see chapter 2). The question we ask is whether the same factors that predict individual differences within the typical range of typically developing children in chapter 2 can also be used to identify late talking children. Late talking children were identified in both samples of children.

In the remainder of this introduction, we outline why the predictors mentioned above are likely candidates for identifying late talking children. This is followed by a summary of the aims of this study.

3.1.1 Demographic factors

Two demographic factors that have been shown to be implicated in language development are child sex and family history of speech or language impairment. Previous research has shown that girls produce their first words and sentences faster than boys (Fenson et al., 1994; Horwitz et al., 2003). In addition, prevalence of late talking has been shown to be greater in boys compared to girls (Rescorla, 2011). However, while being a boy may be a good indicator of late talking, it may not be good at discriminating between late talkers and typically developing children. Thus, it is important to investigate the role that sex plays in language identifying late talkers.

Family history of speech or language impairment has been shown in previous research to predict language development over time (Reilly et al., 2007; Reilly et al., 2009). In addition, family history contributes to language delay, with a greater number of late talking children with a family history of speech or language impairment (Bishop et al., 2012; Rescorla, 2011). However, like child sex, the role of family history in discriminating between typically developing and late talking is less clear cut and therefore, more research is required.

3.1.2 Environmental factors

Caregiver input and conversational turns are robust predictors of individual differences in language acquisition. Both the quality (Cartmill et al., 2013; Demir-Vegter, Aarts, & Kurvers, 2014; Jones & Rowland, 2017) and quantity (Hoff & Naigles, 2002; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991) of caregiver input play an important role in vocabulary development. For example, Cartmill et al. (2013) showed caregiver input at 18 months correlated with children's vocabulary size at 54 months. Furthermore, when combined, the quality and quantity of caregiver input explained 22% of variance in children's vocabulary size.

Similarly, conversational turns are associated with variance in language acquisition with children who engage in more conversational turns achieving higher scores on vocabulary assessments (Romeo et al., 2018; Zimmerman et al., 2009). Zimmerman et al (2013) found that a greater number of conversational turns was associated with an increase in language score; with every 100 fold increase in the number of adult-child conversational turns a child engaged in, their language score increased by 1.92. However, the influence of input and conversational turn on late talking is virtually unexplored. Due to the strength of these factors in predicting variance in vocabulary acquisition, they are likely to be good predictors of late talking and in turn also likely to be good at discriminating between late talkers and typically developing children

3.1.3 Language proficiency factors

Measures of language proficiency have repeatedly been shown to be robust predictors of individual differences in vocabulary acquisition (Henrichs et al., 2011; Rescorla, 2011) and of language delay (Westerlund, Berglund & Eriksson., 2006). For example, Westerlund et al. (2006) found that productive vocabulary at 18 months was the strongest predictor of language delay at 3 years. However, productive vocabulary alone did not yield a high level of sensitivity, only accurately classifying 50% of children with language delay at 3 years. Thus, it is important to investigate how well productive vocabulary, both in isolation and when combined with other factors, can identify late talking children.

However, if children are producing very few words, it may be hard to capture any variance. Therefore, earlier receptive vocabulary may actually be a better predictor of later productive vocabulary than earlier productive vocabulary. The number of words a child knows at an earlier time point has been shown to predict their productive vocabulary size at a later time (Stolt et al., 2016). In addition, lower receptive vocabulary scores predict lower productive vocabulary scores and later language impairment (Hsu & Iyer, 2016). However, there has been little research looking at the role of receptive vocabulary in late talking, and in the research, there is a trend not to control for productive vocabulary size (Paul, Looney, & Dahm, 1991). We predict that receptive vocabulary will be a strong measure of early communicative ability and therefore will explain unique variance in later productive vocabulary as well as being a good predictor of late talking.

Mean length of utterances (MLU) is a measure of language complexity that assesses the average number of morphemes produced in a child's three longest utterances (Brown 1973). The more morphemes produced, the greater the complexity. Individual differences in vocabulary development are associated with individual differences in mean length of utterances (Bates & Goodman, 1997; Rescorla, Dahlsgaard, & Roberts, 2000; Rescorla & Schwartz, 1990). Previous research has shown that vocabulary and grammar develop together with growth in language complexity aiding vocabulary acquisition (Dionne, Dale, Boivin, and Plomin, 2003). Therefore, MLU is likely to be a good predictor of late talking, with delays in vocabulary acquisition also impacting syntax development.

Similarly, gesture skills are a robust predictor of individual differences in productive vocabulary development (Colonnesi, Stams, Koster, & Noom, 2010; Hsu & Iyer, 2016; Iverson & Goldin-Meadow, 2005). Children who produce more gestures elicit more input from their parents and in turn, have more opportunity to learn from this input (Goldin-Meadow, Goodrich, Sauer, & Iverson, 2007; Özçalışkan, Adamson, Dimitrova, & Baumann, 2017). Gesture skills are likely to be a good predictor of late talking; if gestures are a measure of early communicative abilities, children who experience a delay in language acquisition, may also experience a delay in gesture production. In addition, when looking at language skills in very young children, gesture scores may be a good indicator of communitive ability, especially if vocabulary scores are low.

3.1.4 Cognitive factors

Speed of linguistic processing and non-word repetition have both been shown to predict individual differences in productive vocabulary development. Speed of linguistic processing is a measure of a child's reaction time to process familiar linguistic material. It is measured by assessing the length of time it takes for a child to orient their gaze to a named picture (Fernald, Perfors, & Marchman, 2006). Speed of linguistic processing has been shown to be associated with vocabulary development (Fernald & Marchman, 2012; Fernald et al., 2006; Killing & Bishop, 2008; Marchman & Fernald, 2008). For example, Fernald and Marchman (2012) found processing speed at 18 months predicted vocabulary at 30 months. In addition, children whose vocabulary had caught up by 30 months had faster processing speeds at 18 months. Thus, if speed of processing facilitates vocabulary development, slower processing will likely have a negative impact on vocabulary acquisition. Thus, speed of linguistic processing is a promising factor for distinguishing between typically developing and late talking children.

Phonological working memory, as measured by non-word repetition tests, is associated with individual differences in language acquisition (Gathercole & Baddeley, 1989, 1990; Hoff, Core, & Bridges, 2008). Phonological working memory is the ability to temporarily store speech sounds in working memory. In order to learn new words, children need to store the phonological representation of the word in their phonological working memory; and the more speech sounds children can store in their working memory, the more opportunity they have to transfer these into long term memory. Non-word repetition tests, which require children to repeat sequences of nonsense syllables (non-words), assess phonological working memory. The use of non-words means children cannot use their knowledge of words during this task and thus, they must use their phonological working memory to temporarily store and reproduce these words.

Poorer performance on NWR tasks is associated with poorer performance on language assessments, and with language impairment (Conti-Ramsden, Botting, & Faragher, 2001; Thal et al., 2005). For example, Conti-Ramsden et al. (2001) found that NWR correctly identified children with specific language impairment (now developmental language disorder, DLD) with a sensitivity value of up to 78%. However, in this study, and in most others, the children were older than 2 years and thus we are unable to confirm if NWR is equally useful at discriminating between late talking and typically developing children. Studies that have investigated the role of NWR in language acquisition in younger children have not examined the role of NWR had in language delay (Eaton, Newman, Rathner, & Rowe, 2015; Hoff et al., 2008; Roy & Chiat, 2004). For example, Hoff et al. (2008) did find that the NWR performance of children from 20-24 months predicted variance in concurrent vocabulary but did not establish if NWR could be used to identify children whose vocabulary acquisition was delayed. Since NWR is a robust predictor or later language impairment, it is important to investigate its usefulness in identifying early language delay.

The aim of the present study was to investigate whether factors implicated in explaining individual differences in productive vocabulary development at 25 months can also distinguish between typically developing and late talking children. We first ran correlations and regressions to identify factors that predict individual differences in productive vocabulary size at 25 months. We then established the sensitivity and specificity of these factors in identifying late talking toddlers. Finally, we ran discriminant function analyses to establish the discriminatory ability of these factors in distinguishing between typically developing and late talking toddlers. The participants for this study were part of two samples. The first sample were from a longitudinal sample aimed to specifically recruit children whose productive vocabulary was below average from 15-18 months. The second sample were a typically developing comparison sample from the Language 0-5 Project, a longitudinal project run in the North West of England.

3.2 Methods

3.2.1 Participants

The data for this study were collected from two samples of children. One sample of children were part of a project which aimed to recruit children whose productive vocabulary development was below average for their age. The second sample of children were part of the Language 0-5 Project. Details of these two groups are given below.

3.2.1.1 Low language sample

This sample is named the low language sample because only children whose productive vocabulary was below average between 15-18 months were invited to take part in the lab based session at 25 months. The data for the children in this sample was collected from 2017-2019. Parents were invited to complete the UK-CDI, during the recruitment phase, when children were between 15-18 months. In order to specifically recruit children whose language skills were low, all children whose productive vocabulary scores were below the 25th percentile at 15 and 16 months and below the 50th percentile at 17 and 18 months, according to the UK-CDI norms (Alcock, Meints, Rowland, Brelsford, Christopher, & Just, 2020), were invited to the university to take part in two sessions in the lab when they were 25 months old. We chose to use the 50^{th} percentile cut off for 17 and 18 months because the normed productive vocabulary scores are still low at this age (50^{th} percentile at 17 months= 34 words, 50^{th} percentile at 18 months= 44 words) and we did not want to miss children with potential delays by using too strict criteria. A total of 183 families completed the UK-CDI during the recruitment phase; 55 children had vocabulary below the 25th percentile at 15 and 16 months or the 50th percentile at 17 and 18 months. Two of the children were twins so one child was randomly selected to take part. Fifty- four families were invited to participate in the study. Of these, 37 families (69%) agreed to participate in the lab-based study and provided 25 month CDI data. One of these children was bilingual so was excluded from the data analyses. Therefore, a total of 36 children (67% of all invited) were included from this sample. All of these children were mono-lingual English learners. One of the children had ear infections. Details on the number of risk factors is provided in table 4.1.

3.2.1.2 Language 0-5 Project sample

The data for this group were collected as part of the Language 0-5 Project. This project followed children from 6 months to 4 years 6 months from 2014 to 2019. A total of 95 families were recruited (89 at 6 months, six in a second wave at 15 months). Here we use the data collected between 18-25 months. These are the same children whose language was investigated in Jago et al, in submission, see chapter 2. At the time of writing, 17 (18%) of the 95 families had dropped out due to time commitment reasons. Two families had dropped out after the initial sign-up and a further three after the 6 month data-point. Five families dropped out after the 8 months data-point and one after each of the 14, 15 and 19 month data-points. Two more had dropped out after the 21 and 24 month data-points.

We compared children from the Language 0-5 sample and the low language sample on the Bayley Scales of Infant and Toddler Development. This is a standardised measure of child development from 1-42 months. We used the language subscale to obtain standardised scores for receptive and productive vocabulary. The Language 0-5 children completed this when they were 16 months old (mean age =16;29) and the low language sample completed this during the second testing session when they were 25 months old (mean age =25;22). However, because this is a standardised measure, composite scores across different ages are comparable. The mean composite score for the Language 0-5 sample was 111.11 (SD=11.17) and for the low language sample the mean was 101.08 (SD=17.13). All of these children were mono-lingual English learners. None of the children had suffered from any ear infections that could impact language development (e.g. glue ear). See table 4.1 for details on the number of risk factors these children experiences.

We had 25-month productive vocabulary scores available for all 36 children from the low language sample and for 73 children in the Language 0-5 Project sample. For four children in the Language 0-5 Project sample, 25 month productive vocabulary was not available. We used 24 month productive vocabulary as a proxy for 25 month productive vocabulary for these 4 children. Thus, there were 77 children with productive vocabulary scores from the Language 0-5 sample giving a total of 113 children with productive vocabulary scores at 25 months from both samples. For the predictor variables, the total numbers vary because not every child completed every task.

For the descriptive statistics, we split the Language 0-5 sample into two groups; the faster Language 0-5 sample included children whose productive vocabulary was above the 50th percentile at 25 months and the slower Language 0-5 sample included children whose productive vocabulary was below the 50th percentile at 25 months. Therefore, for exploring the data, there are 3 groups: the Faster Language 0-5 sample, the Slower Language 0-5 sample and the low language sample. There were a total of 113 participants. The sample was split almost evenly for sex (girls=59). Most of the children were first borns (n=71) with the remainder having one or more over siblings (n=41).

3.2.2 Overall design

This was a longitudinal study following children from 15 months to 25 months This study was granted ethical approval by The University of Liverpool's Research Ethics Subcommittee for Non-Invasive Procedures for the study Language Development in Late Talkers (Institute Review Board protocol number: RETH000764).

We ran two sets of analyses. In the first set of analyses, we investigated if 10 factors were predictors of individual differences in productive vocabulary at 25 months using correlation and regression analyses. For the regressions, the independent variables were: demographic factors (sex and family history of language delay or dyslexia), environmental factors (adult word count and conversational turn count), language proficiency measures (productive and receptive vocabulary scores recorded between 15-18 months, mean length of utterance of children's three longest utterances (M3L) at 25 months), gesture scores recorded between 15-18 months, and measures of cognitive ability (speed of linguistic processing and non-word repetition (NWR) which measures phonological working memory). The dependent variable was productive vocabulary at 25 months as measured by the Lincoln Communicative Development Inventory (a UK adaptation of the MacArthur Bates CDI Words and Sentences; Meints, Fletcher, & Just, 2017).

In the second set of analyses, we investigated the discriminatory ability of those factors in correctly classifying children late talking and typically developing children. Late talking children were classified in two ways. First, children who were producing fewer than 50 words at 25 months were classified as late talkers (50 word cut off). Second, children who were not combining any words at 25 months were classified as late talking children (failure to combine cut off). We used receiver operating characteristic (ROC) curve analyses to establish the sensitivity and specificity of each factor in correctly classifying late talking and typically developing children. Then we ran discriminant function analyses to identify the accuracy of different combinations of factors in correctly classifying children. The independent variables were six factors that were significant predictors of individual differences in analyses set 1: sex, earlier productive and receptive vocabulary, M3L, earlier gesture scores, and NWR. For the first set of ROC curve and discriminant function analyses, the dependent variable was the first classification of late talking, the 50 word cut off. In the second set of ROC curve and discriminant function analyses, the dependent variable was the second classification of late talking, the failure to combine cut off.

As the data for this study was collected from two samples of children (see below), every effort was made to have data collected at the same age points. There are two exceptions to this. First, earlier productive and receptive vocabulary as well as gesture scores were collected using the UK-CDI from 15-18 months for children in the low language sample. For the Language 0-5 Project sample, we used only data from 18 months. For both samples of children, age adjusted z-scores were used for all analyses (see below for a detailed explanation of this calculation). Second, for language environment measure recordings (LENA) were collected at 25 months for the children in the low language sample and at the 24 month data point for the children in the Language 0-5 Project sample.

3.2.3 Measures

3.2.3.1 Demographic measures

We used the Family Questionnaire to determine sex and family history of language delay or dyslexia for the children in the low language sample between 15-18 months and for children in the Language 0-5 Project when they were recruited at 6 months. The Family Questionnaire is a factual questionnaire that asks about a child's health and family background. This questionnaire was devised for the UK-CDI project (for details of construction, see Alcock et al., 2020). It includes a question about family history of language delay or dyslexia (the question of interest was: "4. Is there anyone in the child's immediate family (brothers/sisters/parents only) with a speech or language difficulty or dyslexia?" (yes/no)). A total of 13 children had a family history of language delay or dyslexia (8 children from the Language 0-5 sample, 5 children from the low language sample).

3.2.3.2 Environmental measures

The Language Environment Analysis (LENA; LENA Research Foundation, 2014) system was used to record audio from the children's environment. One recording was taken for each child at 24 months for the Language 0-5 Project sample (mean= 24 months, 23 days (range=24 months, 6 days- 26 months 1 day) or 25 months for the low language sample (mean= 25 months, 18 days (range25 months, 7 days- 26 months 2 days) months. Recordings were taken for one entire day. Parents were asked to record from morning until their child went to sleep at night during a typical day. We used a LENA Pro digital language processor (DLP; LENA Research Foundation, 2014) to collect these recordings. Children wore these in a specialised t-shirt with a small padded pocket on the front. A LENA Pro DLP can hold up to 16 hours of recording. The mean number of hours recorded was 15 hours, 12 minutes (range: 3hours, 0 minutes- 16hours, 0 minutes).

The LENA Pro processing software (LENA Research Foundation, 2014) automatically analyses the audio recording and provides estimates of adult word count, conversational turn count, and child vocalisation count. We used two measures to assess the language in children's environment: adult word count and conversational turn count. Adult word count is the total number of words spoken by adults within the child's environment. A conversational turn is calculated when a child utterance is followed by an adult utterance, or vice versa, within a 5 second window. We also examined the mean number of child vocalisations for children in each sample. Child vocalisation count is the total number of child vocalisations within the recording. Children in the Language 0-5 sample produced on average 3435 (range=914-7062) vocalisations. Children in the low language sample produced on average 2848 (range=326-5583) vocalisations.

3.2.3.3 CDI questionnaires

Communicative Development Inventories (CDIs) are parent report checklists of words, and gestures or sentences, which require parents to indicate if their child uses, or understands, an item. The UK-CDI Words and Gestures (Alcock et al., 2020) was used to collect vocabulary and gesture scores from 15-18 months (total possible scores are 396 for vocabulary and 57 for gesture). The UK-CDI has been standardised for the UK population, and has good validity and reliability (see Alcock et al., 2020). The Lincoln CDI was used to collect vocabulary scores at 25 months (total possible score = 689) and M3L scores (mean length of the three longest utterances reported by parents (M3L); no upper bound) at 25 months for both groups. For six of the children in the Language 0-5 sample, we did not have M3L scores at 25 months, so M3L at 24 months was used as a proxy for this. The Lincoln CDI is the UK version of the MacArthur Bates CDI Words and Sentences (Meints, et al., 2017). It has not yet been validated but is expected to have good validity and reliability due to its similarity to the MacArthur-Bates instrument. Parents were sent the CDI a week before the lab session and were asked to complete the questionnaire before they come into the university. For both instruments, scores were calculated according to the instructions in the manuals referenced above.

3.2.3.4 Cognitive measures

Speed of Processing

A looking-while-listening paradigm was used to capture children's speed of linguistic processing at 25 months for both groups (note that this differs from Jago et al, in submission, and chapter 2, where speed of processing measures were taken at 19 months). We used an Eyelink ® 1000 Plus eye-tracker (EyeLink 1000 Plus, SR Research Ltd) to capture reaction times (RT) to named familiar words. Children heard a sequence of 64 sentences paired with visual stimuli across two blocks (A and B). Two pictures were shown on the screen paired with a pre-recorded sentence naming one of the pictures (e.g., "Where's the car? Can you see it?"). Words which appeared as the target in block A, appeared as the distractor in block B. Each block also included three filler sentences (e.g., "Do you like the pictures? Here's some more"), to keep the child interested. For 12 sentences (2 in each block) nouns were preceded by an adjective ("blue" or "red"; for example "look at the red ball") to increase the complexity of some sentences (noun-adjective trials). Children were randomly assigned to a counterbalance group (block "A then B" or "B then A").

During the session, parents completed a questionnaire asking whether or not their child understood the target words. Where parents indicated their child did not understand a word, these trials were excluded from any analyses. RT was used to as our measure of speed of linguistic processing. RT was defined as the time taken to initiate a shift from the distractor picture to the target picture between 300ms after and up to 1800ms target word onset. Trials where the child was not looking at the distractor image at the onset of the target word were excluded from the analysis. RT is calculated per trial, per child. A total mean RT for the task was calculated for each child and used for our analyses. A minimum of two trials were required to create a mean score for each child. For the full details of stimuli creation, procedure and coding, see Peter et al. (2019).

The results from the Peter et al. (2019) study found no relationship between speed of processing and vocabulary at 25 months when using the full set of items 64 sentences. The authors proposed that this was due to nouns being high frequency for all children, meaning even slow processors were reacting to them quickly. When the authors compared the relationship between RTs from adjective-noun trials (red ball) and productive vocabulary, the correlations were higher than those using nouns only trials. Therefore, we created two RT scores for each child: RT for all trials and RT for noun-adjective trials.

Non-word repetition

The NWR task used was designed by the Language 0-5 Project team and was completed at 25 months for both groups. Full details of stimuli creation, procedure and coding can be found at The Language 0-5 Project Non-Word Repetition Test site on the Open Science Framework (Bidgood, Durrant, Peter, Pine, & Rowland, unpub., https://osf.io/kau5f/). For this task, children were asked to repeat 24 non-words (6 one-syllable, 6 two-syllable and 6 three-syllable words). At each syllable length, half of the words were wordlike (contained high word average biphone probability) and half of the words were non-wordlike. Words were presented in order of syllable length: 1, 2 then 3 syllables. The task was embedded in a fuzzy felt game in which children attach fuzzy felt animals to a farmyard background. The experimenter produced each non-word a maximum of two times, always with a carrier phrase (e.g. “can you say ...”). The experimenter said the non-word and immediately presented a fuzzy felt picture. When the children attempted the repetition, they were given the picture to attach to the background, regardless of the accuracy of their attempt.

Scores were calculated by totalling the proportion of consonants in each non-word repeated correctly. Consonant errors were coded at the phoneme level. A score of 1 was given for each correct consonant. A score of 0 was given for each omitted or substituted consonant. Added consonants or phonemes and common articulation substitutions were not counted as errors (see Bidgood et al, for details of the coding scheme).

3.2.4 Data coding

3.2.4.1 Classifying late talkers

In the literature, children are classified as late talkers if they are producing fewer than 50 words, or are not combining any words together (Rescorla, 1989). We used both of these criteria for classifying late talkers. First, we classified children as late talkers if they were producing fewer than 50 words (50 word cut off). There were a total of 8 children were classified as late talkers using this cut off. Second, we classified children as late talkers if they were not combining any words by 25 months (failure to combine cut off). We chose to also use this classification because we were interested in how children use language at this age. A total of 14 children were classified late talkers using this cut off.

3.2.4.2 Creation of Z-Scores

The UK-CDI data was collected across different ages from 15-18 months. To control for variances in scores due to the fact that children were different ages when parents completed the UK-CDI, age adjusted z-scores were calculated for each child for productive vocabulary, receptive vocabulary and gestures. We used the mean and standard deviation of productive and receptive vocabulary, and gestures for each age from the entire UK-CDI Project sample (Alcock et al., 2020 to create standardised scores for each child for these three variables. From here, these variables are referred to as earlier productive vocabulary, earlier receptive vocabulary and earlier gesture scores.

3.2.5 Analysis strategy

All analyses were conducted using one-tailed tests as all hypotheses are unidirectional. All outliers were included, unless it was determined that the data point was due to experimenter/participant error (i.e., an impossible value such as a score of 50 on a 40 point scale).

Pearson's correlations were performed using R (R Core Team, 2017, R version 3.4.1) using the R Studio (Version 1.0.153) cor.test function, part of the R base package. Regressions were performed using the lm function from the lme4 package (Bates, Maechler, Bolker & Walker, 2015). For the regression analyses, all categorical variables were sum-coded to make the parameter estimates more straightforward to interpret and all numeric variables were centred.

Receiver operating characteristic (ROC) curves were constructed in SPSS Statistics 24 (Hajian-Tilaki, 2013) to determine the sensitivity and specificity of each of our factors at predicting late talking status. Next we took the cut-off values established in the ROC curve analyses and ran discriminant function analyses in SPSS Statistics 24 to determine the power of different combinations in predicting late talking status at 25 months. We used two different combinations of variables: 1) all of the variables which are successful in predicting unique variance in the regression analyses; 2) only non-experimental measures from earlier time points. We chose these combinations to first identify the discriminatory power of all variables combined, then to determine if earlier, non-experimental measures could be as accurate at predicting late talking status at 25 months.

3.3 Results

In this study, we wanted to: (1) assess the ability of our predictor variables in predicting individual differences in productive vocabulary at 25 months, and (2) investigate the discriminatory ability of these variables in distinguishing between late talking and typically developing children.

We chose not to apply corrections for multiple testing because we did not want to miss factors which could be predictive of individual differences or discriminate between late talking and typically developing children as a result of applying a conservative criterion.

3.3.1 Descriptive statistics

Table 3.1 provides details about demographic factors for participants in both the Language 0-5 sample and the low language sample. To help with the comparison between the two samples, the Language 0-5 sample have been split into two groups, children whose productive vocabulary was above the 50^{th} percentile at 18 months (the faster Language 0-5 group) and children whose productive vocabulary was below the 50^{th} percentile at 18 months (the slower Language 0-5 group). For one participant in the Language 0-5 Project sample, 18 month UK-CDI data was not available. We therefore used their UK-CDI data from 16 months to establish which group they were in. This data was used only for calculating the rate of demographic risk factors of children in each group (table 3.1) and was not used for any further analyses. This table shows that there were more boys in both the low language sample and the slower Language 0-5 sample compared to the faster Language 0-5 sample. Rate of family history of language delay or dyslexia was not different between the three groups. Low socioeconomic status did differ across the groups. Rate of health risk factors (prematurity, low birth weight and reports of ear infections or problems with hearing) were greater in the low language sample. Table 3.2 provides descriptive statistics for each predictor variable for the three groups including number of participants, mean age and age range of when they completed each measure.

Figure 1 shows productive vocabulary scores at 25 months split between the Language 0-5 sample and the low language sample. This graph shows that there is a trend for children in the low language sample (i.e. the children who had low productive vocabulary scores at 15-18 months) to still have productive vocabulary scores lower than those of children in the Language 0-5 sample at 25 months.

Table 3.1:

Number of children experiencing demographic risk factors split by three groups: faster language 0-5 group, slower language 0-5 group, and the low language sample

	Group								
	Faster Language 0-5 sample			Slower Language 0-5 sample			Low language sample		
Variable (risk factor)	*n* (girls)	*n* with risk factor	Mean age tested (y;m;d) (range)	*n* (girls)	*n* with risk factor	Mean age tested: (y;m;d) (range)	*n* (girls)	*n* with risk factor	Mean age tested: (y;m;d) (range)
Sex (being a boy)	49 (32)	17	0;7.10(05;24-1;3;19)	28(7)	21	0;6;26(0;6;00-1;3;00)	36(15)	21	1;4;24(1;3;01-1;6;21)
Family history (yes)	49 (32)	6	0;7;10(0;5;24-1;3;19)	28(7)	2	0;6;26(0;6;00-1;3;00)	36(15)	5	1;4;24(1;3;01-1;6;21)
Socioeconomic status (low)	49 (32)	4	0;7;10(0;5;24-1;3;19)	28(7)	0	0;6;26(0;6;00-1;3;00)	36(15)	4	1;4;24(1;35;01-1;6;21)
Prematurity (born before 37 wks)	49 (32)	0	07;10(05;24-1;3;19)	28(7)	0	06;26(06;00-1;3;00)	36(15)	3	1;4;24(1;3;01-1;6;21)
Low birth weight (below 5lbs 9oz)	49 (32)	0	07;10(05;24-1;3;19)	28(7)	0	06;26(06;00-1;3;00)	36(15)	2	1;4;24(1;3;01-1;6;21)
Ear problems (yes)	49 (32)	0	0;7;10(0;5;24-1;;19)	28(7)	0	0;6;26(0;6;00-1;3;00)	36(15)	1	1;4;24(1;3;01-1;6;21)

Table 3.2*:*

Number of children, mean age and age range for each predictor variable split by sex and by three groups: faster language 0-5 group, slower language 0-5 group, and the low language sample

Girls		Group							
		Faster Language 0-5 sample			Slower Language 0-5 sample			Low language sample	
Variable	n (%)	Mean Age; y;m;d (range)	Mean score (range)	n (%)	Mean Age; y;m;d (range)	Mean score (range)	n (%)	Mean Age; y;m;d (range)	Mean score (range)
Adult word count	31 (66%)	2;0;23 (2;0;8-2;1;25)	19592.10 (4343.00-36862.00)	7 (27%)	2;0;29 (2;0;12-2;1;18)	24227.86 (9319.00-53588.00)	13 (41%)	2;1;16 (2;1;9-2;1;25)	18923.77 (3481.00-32745.00)
Conversational turn count	31 (66%)	2;0;23 (2;0;8-2;1;25)	950.45 (255.00-1579.00)	7 (27%)	2;0;29 (2;0;12-2;1;18)	1029.43 (522.00-1330.00)	13 (41%)	2;1;16 (2;1;9-2;1;25)	818.69 (209.00-1354.00)
Earlier productive vocabulary	32 (67%)	1;6;16 (1;6;0-1;6;29)	0.34 (-0.48-2.65)	6 (22%)	1;6;18 (1;6;7-1;6;29)	-0.73 (-0.94--0.67)	15 (42%)	1;4;21 (1;3;1-1;6;15)	-0.69 (-0.87--0.38)
Earlier receptive vocabulary	32 (67%)	1;6;16 (1;6;0-1;6;29)	0.33 (-1.21-1.68)	6 (22%)	1;6;18 (1;6;7-1;6;29)	-0.30 (-1.01-0.24)	15 (42%)	1;4;21 (1;3;1-1;6;15)	-0.81 (-1.73-0.12)
25 month productive vocabulary	32 (65%)	2;1;10 (2;0;15-2;1;27)	495.78 (192.00-673.00)	7 (26%)	2;1;12 (2;1;6-2;1;23)	388.86 (174.00-503.00)	15 (42%)	2;1;6 (2;1;0-2;1;20)	233.60 (41.00-509.00)
M3L at 25 months	32 (65%)	2;1;10 (2;0;15-2;1;27)	5.67 (2.67-14.67)	7 (26%)	2;1;12 (2;1;6-2;1;23)	4.10 (1.00-8.33)	15 (42%)	2;1;6 (2;1;0-2;1;20)	3.09 (1.00-5.67)
Earlier gestures	32 (67%)	1;6;16 (1;6;0-1;6;29)	0.94 (-0.52-1.88)	6 (22%)	1;6;18 (1;6;7-1;6;29)	0.68 (0.31-1.46)	15 (42%)	1;4;21 (1;3;1-1;6;15)	-0.23 (-1.29-1.30)
Speed of linguistic processing	29 (67%)	2;1;13 (2;1;2-2;1;28)	671.52 (438.05-1080.33)	6 (24%)	2;1;13 (2;1;6-2;1;23)	692.54 (500.75-812.33)	13 (43%)	2;1;10 (2;1;2-2;1;21)	756.92 (611.89-1056.20)
NWR	26 (62%)	2;1;12 (2;1;0-2;1;28)	0.73 (0.45-0.96)	7 (33%)	2;1;13 (2;1;6-2;1;23)	0.62 (0.37-0.81)	9 (43%)	2;1;11 (2;1;2-2;1;21)	0.46 (0.32-0.69)

Table 3.2 continued:

Number of children, mean age and age range for each predictor variable split by sex and by three groups: faster language 0-5 group, slower language 0-5 group, and the low language sample

Boys	Group								
	Faster Language 0-5 sample			Slower Language 0-5 sample			Low language sample		
Variable	n (%)	Mean Age; y;m;d (range)	Mean score (range)	n (%)	Mean Age; y;m;d (range)	Mean score (range)	n (%)	Mean Age; y;m;d (range)	Mean score (range)
Adult word count	16 (34%)	2;0;23 (2;0;7-2;1;15)	17983.06 (7639.00-28738.00)	19 (73%)	2;0;20 (2;0;6-2;1;4)	19383.42 (8578.00-34310.00)	19 (59%)	2;1;19 (2;1;7-2;2;2)	17786.90 (6819.00-42451.00)
Conversational turn count	16 (34%)	2;0;23 (2;0;7-2;1;15)	947.94 (447.00-1406.00)	19 (73%)	2;0;20 (2;0;6-2;1;4)	925.68 (330.00-2069.00)	19 (59%)	2;1;19 (2;1;7-2;2;2)	887.90 (119.00-2126.00)
Earlier productive vocabulary	16 (33%)	1;6;18 (1;6;12-1;6;28)	0.12 (-0.43-1.28)	21 (78%)	1;6;19 (1;6;4-1;7;3)	-0.72 (-0.99--0.51)	21 (58%)	1;4;27 (1;3;5-1;6;21)	-0.77 (-1.00--0.58)
Earlier receptive vocabulary	16 (33%)	1;6;18 (1;6;12-1;6;28)	0.11 (-1.51-1.60)	21 (78%)	1;6;19 (1;6;4-1;7;3)	-0.87 (-2.19-1.54)	21 (58%)	1;4;27 (1;3;5-1;6;21)	-0.82 (-2.38-2.04)
25 month productive vocabulary	17 (35%)	2;1;15 (2;0;24-2;2;1)	473.65 (351.00-630.00)	21 (78%)	2;1;12 (2;0;7-2;1;27)	254.76 (13.00-574.00)	21 (58%)	2;1;10 (2;0;14-2;2;0)	156.14 (1.00-552.00)
M3L at 25 months	17 (35%)	2;1;15 (2;0;24-2;2;1)	5.09 (3.00-8.00)	21 (78%)	2;1;12 (2;0;7-2;1;27)	3.23 (1.00-10.00)	21 (58%)	2;1;10 (2;0;14-2;2;0)	2.55 (1.00-6.33)
Earlier gestures	16 (33%)	1;6;18 (1;6;12-1;6;28)	0.69 (-0.62-1.88)	21 (78%)	1;6;19 (1;6;4-1;7;3)	0.02 (-2.08-1.77)	21 (58%)	1;4;27 (1;3;5-1;6;21)	-0.46 (-1.86-1.58)
Speed of linguistic processing	15 (35%)	2;1;14 (2;0;24-2;1;28)	683.32 (502.67-904.57)	20 (80%)	2;1;14 (2;1;0-2;2;2)	680.26 (521.75-947.33)	17 (57%)	2;1;14 (2;1;5-2;2;0)	725.67 (568.06-897.47)
NWR	16 (38%)	2;1;15 (2;0;24-2;2;10)	0.68 (0.28-0.89)	15 (71%)	2;1;17 (2;1;0-2;2;2)	0.49 (0.08-0.78)	12 (57%)	2;1;16 (2;1;5-2;2;0)	0.44 (0.25-0.81)

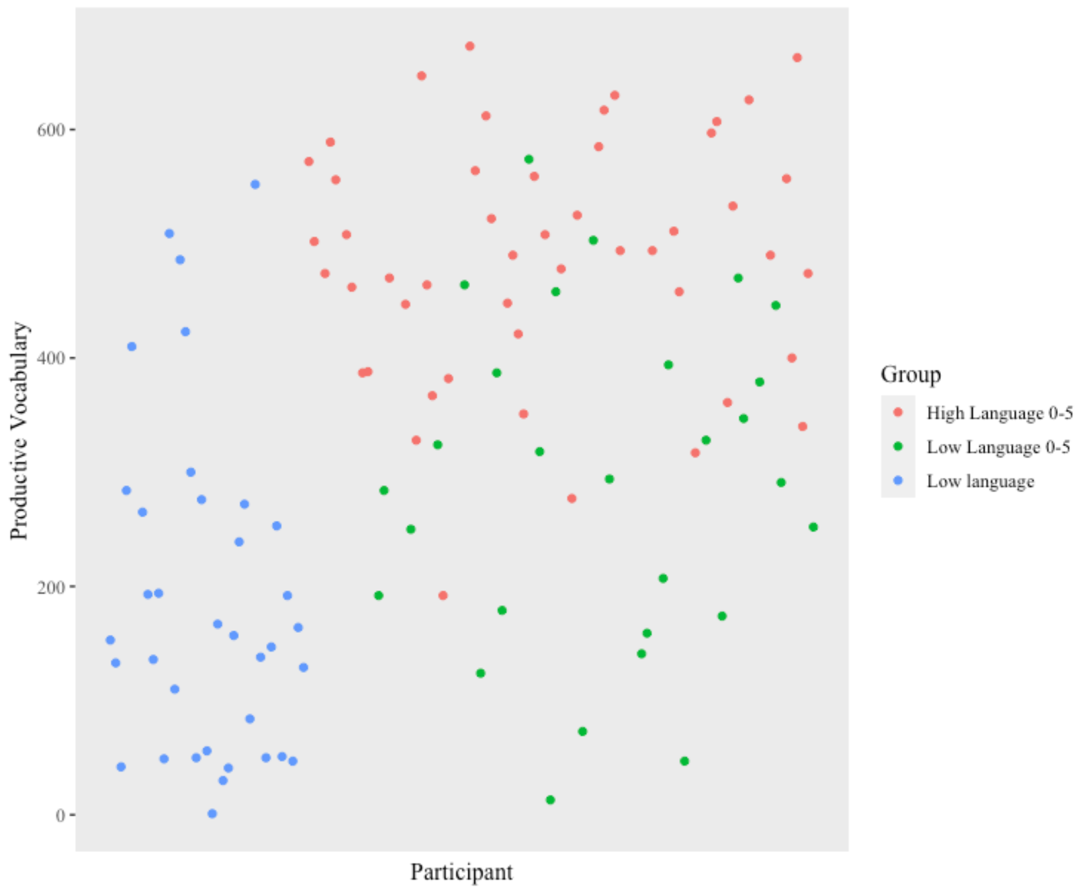

Figure 3.1: productive vocabulary scores for children in the Language 0-5 sample and the low language sample *variable split by three groups: faster Language 0-5 group, slower Language 0-5 group, and the low language sample*

Of the low language sample, only children whose productive vocabulary was below the 25th percentile at 15 and 16 months and below the 50th percentile at 17 and 18 months were invited to take part in the lab-based. Details of mean productive vocabulary, range of vocabulary scores and number of children who were invited for the follow-up project and children who were not invited for the follow-up can be seen in table 3.3.

Table 3.3

Mean productive vocabulary scores, range of vocabulary scores and n for children who were and were not invited for the follow-up project. Mean and range are given in raw scores and standardised z-scores

Month	Group	*M* (z-score)	Range (z-score)	*n* (girls)
15	Invited	4.70 (-0.81)	0-8 (-0.95- -0.70)	10(6)
	Not invited	22.19 (-0.27)	9-57 (-0.67-0.80)	31(16)
16	Invited	4.42 (-0.80)	0-10 (-0.87- -0.70)	12(4)
	Not invited	47.30 (-0.07)	14-174 (-0.64-2.06)	30(19)
17	Invited	12.46 (-0.65)	0-34 (-0.81- -0.38)	26(7)
	Not invited	86.89 (0.29)	39-228 (-0.31-2.07)	29(13)
18	Invited	15.43 (-0.83)	0-43 (-1.00- -0.52)	7(3)
	Not invited	109.24 (0.23)	49-320 (-0.45-2.60)	17(9)

3.3.2 Predicting individual difference in productive vocabulary at 25 months

We ran this analysis in two steps. In step 1, we assessed the relationship between our predictor variables and the children's 25 month productive vocabulary. We did this to identify which variables were related to 25 month productive vocabulary, to retain them for the remaining analyses. For categorical factors we ran simple linear regressions and for numeric factors we ran correlations. In step 2, once we identified factors that were significantly related to 25 month productive vocabulary, we entered these into multiple regressions to determine if they added additional explanatory power after controlling for earlier productive vocabulary and sex.

The results of the simple linear regressions can be seen in table 3.4. Child sex was a significant predictor of 25 month productive vocabulary, explaining 12% of the variance in scores. As girls were coded as 1 and boys were coded as -1, the positive beta value means that being a girl is associated with higher productive vocabulary scores. Therefore, sex was retained for the regression analyses. Family history of language delay or dyslexia was not a significant predictor of productive vocabulary at 25 months and was not retained for the regression analyses.

The results of the correlation analyses can be seen in table 3.4. Neither of our environmental variables (adult word count and conversational turn count) correlated significantly with productive vocabulary at 25 months and therefore, were not retained for the regression analyses. All three language variables (earlier productive and receptive vocabulary, and M3L) and gestures significantly correlated with productive vocabulary at 25 months, and were retained for the regression analyses. Of our cognitive variables, NWR and speed of linguistic processing on adjective-noun trials significantly correlated with productive vocabulary at 25 months and so were retained for the regression analyses. Speed of linguistic processing on all trials did not correlate significantly and therefore, was not retained for the regression analyses.

Table 3.4

Relationship between predictor variables and 25 month productive vocabulary scores. Statistics given are results of linear regression for categorical predictor variables and correlations for numeric variables

Category	Variable	R^2	*SE*	*t*	*r*	*p*
Demographics	Sex	.12	16.42	3.85		<.001
	Family history	.001	27.34	0.35		.73
Environmental variables	Adult word count				.06	.55
	Conversational turn count				.11	.27
Language variables (speech)	Earlier productive vocabulary				.65	<.001
	Earlier receptive vocabulary				.58	<.001
	25 month M3L				.65	<.001
Gestures	Earlier gestures				.56	1.69e-10
Cognitive variables	Speed of linguistic processing				-.06	.53
	Speed of linguistic processing on adjective-noun trails				-.36	.001
	NWR				.73	<.001

To check if any of the predictor variables were highly collinear, we ran correlation analyses between each predictor variable. None of the variables correlated highly, with all below r=.80, so all were kept for the regression analyses.

In step 2, we ran regression using all the variables retained from step 1 (sex, earlier productive and receptive vocabulary, M3L, earlier gestures, speed of processing on adjective-noun trials and NWR). We were interested in seeing if any of the predictor variables remained significant predictors after controlling for sex and earlier productive vocabulary, which have been shown, in previous research, to be robust predictors of vocabulary size. Thus, we controlled for sex and earlier productive vocabulary by entering them first into our base model. For each model after, we individually entered each additional predictor variable that we retained from step 1, on top of the base model.

The results from the base model can be seen in table 3.5 and the results for the additional five regressions can be seen in table 3.6. Again, we chose not to correct for multiple comparisons in order to avoid removing potentially influential variables for our sensitivity and specificity analysis below, but note that applying a conservative Bonferroni correction would reduce the critical p value to 0.008.

Table 3.5

Base multiple regression model with sex and 18 month productive vocabulary as predictors and 24M productive vocabulary size as the outcome measure.

Base Model	R^2	Adj. R^2	*F*	*df*	*p*
Sex + Earlier productive vocabulary	.45	.44	44.38	2,109	<.001
		B	*SE*	*T*	*p*
Sex		32.41	13.74	2.36	.02
Earlier productive vocabulary		154.74	19.36	7.99	<.001

Note. Adj. =Adjusted

The base model explains a significant proportion of the variance in productive vocabulary scores at 25 months (45%). Additional significant variance was explained by both language variables (earlier receptive vocabulary explained an additional 6% of the variance, M3L explained an additional 16% of the variance), earlier gesture scores (explained an additional 5% of the variance) and NWR (explained an additional 18% of the variance). Speed of linguistic processing on adjective-noun only trials did not explain any additional unique variance in productive vocabulary at 25 months.

Table 3.6

Results of six separate multiple regression models. Each model contains the baseline predictor variables plus one additional predictor variable, specified below

Model	Variables	R^2	Adj. R^2	*F*	*df*	*p*
Base model+ earlier receptive vocabulary		.51	.49	37.12	3,108	<.001
			B	*SE*	*T*	*p*
	Sex		27.55	13.12	2.10	.04
	Earlier productive vocabulary		113.93	21.61	5.27	<.001
	Earlier receptive vocabulary		53.27	14.83	3.59	<.001
Base model+ 25 month M3L		R^2	Adj. R^2	*F*	*df*	*p*
		.61	.60	56.06	3,106	<.001
			B	*SE*	*T*	*p*
	Sex		25.43	11.83	2.15	.03
	Earlier productive vocabulary		83.58	19.66	4.25	<.001
	M3L		38.24	5.89	6.49	<.001
Base model+ earlier gesture scores		R^2	Adj. R^2	*F*	*df*	*p*
		.50	.49	36.45	3,108	<.001
			B	*SE*	*T*	*p*
	Sex		24.69	13.30	1.86	.07
	Earlier productive vocabulary		120.63	20.97	5.75	<.001
	Earlier gestures		56.11	16.35	3.43	<.001
Base model+ speed of linguistic processing on adjective-noun trials		R^2	Adj. R^2	*F*	*df*	*p*
		.45	.43	19.84	3,73	<.001
			B	*SE*	*T*	*p*
	Sex		32.92	14.67	2.24	.03
	Earlier productive vocabulary		103.91	21.43	4.84	<.001
	Speed of linguistic processing on adjective-noun trials		-0.09	0.06	-1.40	.16;
Base model+ NWR		R^2	Adj. R^2	*F*	*df*	*p*
		.63	.61	44.96	3,80	<.001
			B	*SE*	*T*	*p*
	Sex		15.20	11.56	1.31	.19
	Earlier productive vocabulary		71.05	16.91	4.20	<.001
	NWR		431.58	69.80	6.18	<.001

Note: Adj. =Adjusted

We then ran one final regression which included all of the variables which were significant predictors in the last set of regressions. This full model explained 75% of the variance in productive vocabulary scores at 25 months. In addition, earlier receptive vocabulary, M3L and NWR all remained significant unique predictors. The results of this full model can be seen in table 3.7.

Table 3.7

Full regression model predicting variance in 25 month productive vocabulary scores including all significant predictors from the last set of regression analyses

Model	Variables	R^2	Adj. R^2	*F*	*df*	*p*
Full model		.75	.73	37.41	6,75	<.001
			B	*SE*	*T*	*p*
	Sex		13.77	9.94	1.39	.17
	Earlier productive vocabulary		16.17	17.42	0.93	.36
	Earlier receptive vocabulary		33.42	12.46	2.68	.009
	M3L		21.65	5.22	4.15	<.001
	Earlier gesture score		15.53	13.63	1.14	.26
	Non-word repetition		297.72	63.67	4.68	<.001

Adj. =Adjusted

In order to understand why speed of processing using adjective-noun trials fails to explain additional variance in productive vocabulary after controlling for earlier productive vocabulary and child sex, we examined the mean and SD of RT scores obtained by children in the three groups created above: faster Language 0-5 sample, slower Language 0-5 sample and low language sample. Children in the faster Language 0-5 sample had a mean RT of 760.60 (SD=259.51). Children in the slower Language 0-5 sample had a mean RT of 870.00 (SD=190.10). Children in the low language sample had a mean RT of 1084.31 (SD=184.80). Thus, although the lower language sample had slower speed of processing, as predicted, there was substantial overlap (large SDs), between the scores in the three groups, especially between children in the two Language05 groups.

3.3.3 Classifying late talking children

The aim of these analyses was to determine how accurately each of our predictor variables classified children who were and were not late talkers. The independent variables were those that significantly predicted variance in productive vocabulary in our analyses above: sex, earlier productive and receptive vocabulary, M3L, earlier gestures and NWR. The dependent variable for the first set of analyses was late talking status as defined by the *50 word* cut off. The dependent variable for the second set of analyses was late talking status as defined by the *failure to combine* cut off. As ROC curve analyses measure the accuracy of a continuous measure in predicting a binary outcome, sex could not be included in the ROC curve analyses but was reintroduced in the discriminant function analyses below.

First, we ran ROC curve analyses to determine the diagnostic accuracy of each of the predictor variables. This analysis does this by measuring the overlap in scores obtained for each variable by the two groups of children. An ROC curve is created by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) at different thresholds. This yields an area under the curve (AUC) which is a measure of the overall accuracy of each predictor in correctly classifying children into the two groups. AUC values range from 0.5-1.0; higher AUC values indicate better accuracy. This analyses also provides levels of sensitivity and specificity for different scores for each variable. Here, the sensitivity indicates the ability of a predictor to correctly classify late talking children (the true positive rate). Specificity indicates the ability of a predictor to correctly classify typically developing children (the true negative rate). Sensitivity and specificity rates between 70-80% are considered good for diagnostic assessments used for autistic spectrum disorder (Bright Futures Steering Committee & Medical Home Initiatives for Children with Special Needs Project Advisory Committee, 2006). Therefore, we considered values above 70% to be good in the present analyses.

Results from the first five ROC curve analyses can be seen in table 3.8 (classifying late talking using the 50 word cut off). This table also provide the best cut-off scores in terms of maximising both sensitivity and specificity. Earlier productive vocabulary, M3L and NWR were successful in correctly classifying late talking and typically developing children with high AUC values (0.86, 0.91 and 0.99 respectively) as well as yielding good sensitivity and specificity. While the AUC for earlier gestures was good (0.71) the sensitivity of this measure was poor (63%). Therefore, earlier gesture scores are less accurate in identifying late talking children. For earlier receptive vocabulary, the AUC was poorer, 0.68. This is reflected in the specificity of this model, 53%, meaning this measure was less accurate in identifying typically developing children.

Table 3.8

ROC Analyses Results: Diagnostic classification accuracy for each predictor variable correctly distinguishing between typically developing and late talking children using the 50 word cut off

Variable	Best cut-off (raw scores)	AUC	SE	*p*	Sensitivity	Specificity
Earlier productive vocabulary	≤-0.73 words (≤ 24 words)	0.86	0.04	<.001	0.88	0.76
Earlier receptive vocabulary	≤-0.53 words (≤ 207 words)	0.68	0.09	.088	0.88	0.53
M3L	≤ 2.83 morphemes	0.91	0.03	<.001	0.88	0.81
Earlier gesture scores	≤-0.36 gestures (≤ 42 gestures)	0.71	0.10	.049	0.63	0.76
NWR	≤26% accuracy	0.99	0.01	.019	1.00	0.99

Note: earlier vocabulary and gesture scores are age-standardised scores. Raw scores are given in brackets beside each standardised score.

Results from the second four ROC curve analyses can be seen in table 3.9 (classifying late talking children using the failure to combine cut off). As the dependent variable was whether or not children were combining words together, M3L was not included in this set of because it was measured at 25 months, so including it as an independent variable would be too circular. The AUC for all models was good. Both earlier gesture scores and NWR were successful in correctly classifying late talking and typically developing children, yielding good sensitivity and specificity (earlier gestures: sensitivity 77%, specificity 72%; NWR: sensitivity 100%, specificity 86%). Note that 100% sensitivity for the NWR measures means that this model correctly classified all seven late talking children. While the AUC values for earlier productive and receptive vocabulary were good (0.79 and 0.75 respectively) the sensitivity of these measures was poorer (earlier productive vocabulary: sensitivity 69%, specificity 77%; earlier receptive vocabulary:

sensitivity 69%, specificity 77%) meaning these two variables did not do well in correctly classifying late talking children.

Table 3.9

ROC Analyses Results: Diagnostic classification accuracy for each predictor variable correctly distinguishing between typically developing and late talking children using failure to combine words as the cut-off for late talking

Variable	Best cut-off (raw score)	AUC	SE	*p*	Sensitivity	Specificity
Earlier productive vocabulary	≤ -0.73 words (≤ 24 words)	0.79	0.06	<.001	0.69	0.77
Earlier receptive vocabulary	≤-1.07 words (≤ 161 words)	0.75	0.08	.004	0.69	0.77
Earlier gesture scores	≤-0.15 gestures (≤ 44 gestures)	0.82	0.06	<.001	0.77	0.72
NWR	≤41% accuracy	0.93	0.03	<.001	1.00	0.86

Note: earlier vocabulary and gesture raw scores are standardised scores converted back to raw scores at 18 months

We ran one final exploratory ROC curve analyses to investigate the discriminatory ability of speed of processing on adjective-noun trials. While this variable did not predict unique variance in productive vocabulary, previous research has shown that speed of linguistic processing is related to syntax acquisition and thus, it is possible that it would be a good discriminator between late talking and typically developing children. We used the failure to combine classification because if speed of processing as related to syntax development, it seems logical that our classification be based on syntax. See table 3.10 for the results of this analysis. Speed of linguistic processing on adjective-noun trials did not successfully classify late talking and typically developing children. The AUC was below the threshold, and both the sensitivity (60%) and specificity (60%) were poor. Note, there were only five late talking children with speed of processing available for adjective-noun trials.

Table 3.10

ROC Analyses Results: Diagnostic classification accuracy for speed of processing on adjective-noun trials correctly distinguishing between typically developing and late talking children using failure to combine words as the cut-off for late talking

Variable	Best cut-off	AUC	SE	*p*	Sensitivity	Specificity
Speed of processing on adjective-noun trials	≥930.33	0.61	0.15	.43	0.60	0.60

Next, we ran two different discriminant function analyses, to determine if different combinations of our predictor variables were better at classifying late talking children compared to each predictor individually. We ran these two analyses twice; first using the 50 word cut off classification and then using the failure to combine classification. We recoded our predictor variables into categorical variables (1, 2) based on the best cut-off scores from the ROC curve analyses above. For example, the best cut off score for earlier productive vocabulary was *producing -.73 or fewer words* using the age adjusted standardised scores. Thus, we created a new variable where children who were producing -.73 or fewer words were coded as 1 and children producing more than -.73 words were coded as 2. This was done for both sets of classifications using the cut off scores from the associated ROC curve analysis. The results from all discriminant function analyses can be seen in tables 3.11 and 3.12. As with the ROC curve analyses above, M3L was not included for the analyses using the failure to combine classification. See table 3.13 for the standardised canonical discriminant function coefficients. These values provide the contribution of each variable in classifying children into their groups (late talking and typically developing).

The first discriminant function analysis included all of the variables that were included in the ROC curve analyses as well as sex. For the 50 word classification criteria, this model correctly classified late talking and typically developing children with an accuracy of 98.80%. While the sensitivity and specificity of this model are good (100.00% sensitivity and 98.80% specificity), note that there were only two late talking children available for this analysis (i.e. who had data for all of these variables). For the failure to combine classification criteria, this model correctly classified late talking and typically developing children with an accuracy of 91.80%.

Both the sensitivity and specificity of this model were good (83.30% sensitivity and 92.40% specificity).

The second discriminant function analysis included only non-experimental variables from earlier time points to determine if we can identify late talking status from an earlier age: sex, earlier productive and receptive vocabulary, and earlier gesture scores. When using the 50 words classification, this model correctly classified late talking and typically developing children with an accuracy of 92.80%. However, the sensitivity of this model was 0.00%, meaning it did not correctly classify any late talking children. When using the failure to combine classification, this model correctly classified children with an accuracy of 87.70%. Using this cut off, increased the sensitivity of this model to 46.20%. This model is also successful in classifying typically developing children with a specificity of 93.10%.

We see from these analyses that M3L and NWR are consistently the best predictors of group membership using the 50 word cut off, and earlier gestures and NWR are the best predictors when using the failure to combine cut off. Therefore, we ran another set of discriminant function analyses using only the best predictors. For the 50 word classification criteria of late talking, this model yielded good accuracy, 98.80%. Both the sensitivity (100.00%) and specificity (98.80%) of this model were good. However, as with the first model, there were only two late talking children included in this model. When using the failure to combine cut off, this model yielded good accuracy, 87.20%. The sensitivity and specificity of this model were also good, 100.00% and 86.30% respectfully.

Table 3.11

Results of the discriminate function analyses for various combinations of variables for predicting late talking status as determined by producing fewer than 50 words at 25 months

Variable	*r*	χ^2	*n* (late talkers)	*df*	*p*	Accuracy	Sensitivity	Specificity
All variables (sex, earlier productive vocabulary, earlier receptive vocabulary, M3L, earlier gestures and NWR)	.81	80.50	80 (2)	6	<.001	98.80%	100.00%	98.70%
No experimental data from earlier time points (sex, earlier productive vocabulary, earlier receptive vocabulary, earlier gestures)	.38	17.08	111 (8)	4	.002	92.80%	0.00%	100.00%
Best predictors (M3L, NWR)	.81	85.94	81 (2)	2	<.001	98.80%	100.00%	98.80%

Table 3.12

Results of the discriminate function analyses for various combinations of variables for predicting late talking status as determined by failure to combine together words at 25 months

Variable	*r*	χ^2	*n* (late talkers)	*df*	*p*	Accuracy	Sensitivity	Specificity
All variables (sex, earlier productive vocabulary, earlier receptive vocabulary, earlier gestures and NWR)	.60	35.22	85 (6)	5	<.001	91.80%	83.30%	92.40%
No experimental data from earlier time points (sex, earlier productive vocabulary, earlier receptive vocabulary, earlier gestures)	.43	22.58	114 (13)	4	<.001	87.70%	46.20%	93.10%
Best predictors (earlier gestures, NWR)	.56	30.50	86 (6)	2	<.001	87.20%	100.00%	86.30%

Table 3.13

Standardised canonical coefficients for all variables in the two different discriminant function analyses predicting late talking status at 25 months

Model	Variable	*r* (producing fewer than 50 words)	*r* (not combining words)
All variables	Sex	0.00	0.04
	Earlier productive vocabulary	0.00	0.15
	Earlier receptive vocabulary	0.00	0.42
	M3L	0.00	*NA*
	Earlier gestures	0.00	-0.35
	NWR	1.00	0.93
Non-experimental	Sex	0.29	0.30
	Earlier productive vocabulary	0.77	0.40
	Earlier receptive vocabulary	0.14	0.36
	Earlier gestures	0.12	0.39
Best predictors	Earlier gestures	*NA*	-0.13
	M3L	0.00	*NA*
	NWR	1.00	1.03

3.4 Discussion

The goal of the present study was to investigate whether predictors of individual differences in productive vocabulary development up to 25 months, could also discriminate between typically developing and late talking children in a sample of 113 children. First, we investigated whether our measures predicted individual differences in productive vocabulary at 25 months. In line with our predictions, all language measures (earlier productive and earlier receptive vocabulary as well as M3L), gestures, and two cognitive measures (speed of linguistic processing on adjective-noun trials, and NWR) correlated with productive vocabulary. We were interested in whether our measures predicted unique variance in productive vocabulary at 25 months after controlling for child sex and earlier productive vocabulary (recorded between 15-18 months). Earlier receptive vocabulary, M3L, earlier gestures and NWR all remained unique predictors of productive vocabulary at 25 months when controlling for earlier productive vocabulary and child sex.

Our first null finding can be easily explained. Family history of speech or language impairment has been previously shown to be a robust predictor of individual differences in vocabulary development (Reilly et al., 2007; Reilly et al., 2009). The failure to find a relationship between family history and productive vocabulary at 25 months is likely a result of the low number of children with any family history of speech or language impairment in the entire sample. There were only 13 out of 113 children with a family history of speech or language impairment and therefore, there was not enough variance to capture an effect.

For the two environmental factors (adult word count and conversational turn count), one of the null findings can be easily explained. Adult word count, as measured by LENA, captures both overheard speech and child directed speech. Previous research has shown that child direct speech is more important than overheard speech for children's vocabulary development (Weisleder & Fernald, 2013). The finding that conversational turns did not correlate with vocabulary at 25 months is more surprising, especially given that in a previous paper using the data from the children in the Language 0-5 Project sample (chapter 2, Jago et al, in submission), conversational turn count was a significant predictor of productive vocabulary at 24 months. One possible explanation for this finding is that while conversational turns are predictive of vocabulary development in typically developing children (Romeo et al., 2018; Zimmerman et al., 2009), it is not a predictor of late talking status; i.e. that late talking children's delay is caused by something other than quality or quantity of their input. Thus, the inclusion of late talking children in this sample may have impacted the relationship between conversational turns and vocabulary. Previous research looking at the role of conversational turns in language development has not investigated its role in language delay (Gilkerson et al., 2017; Romeo et al., 2018; Zimmerman et al., 2009).

The finding that speed of processing, on all trials (adjective-noun and noun only trials) did not correlate with vocabulary at 25 months is consistent with previous research (Peter et al., 2019). This is likely a result of the items used in the task; if the target nouns were too high frequency for all children, there would not be much variance in the scores obtained and thus reduce the chances of finding an effect. This idea is supported by the results from the speed of processing on adjective-noun only trials. Here we found a relationship with 25 month productive vocabulary. The introduction of adjectives created variance in RT scores. This is likely due to the adjective target words being lower frequency for some children.

However, after controlling for child sex and earlier productive vocabulary, this effect disappeared; speed of processing explained no additional unique variance. When we compare the mean RTs of children, divided by three groups we find again that there is little variance in scores obtained between children in both Language 0-5 groups (of the 77 children with speed of processing scores on adjective-noun trials, 61 of them were in these two groups). Research by Fernald et al. (2006) proposes that speed linguistic processing would predict unique variance in vocabulary because speed of processing contributes to vocabulary growth over and above concurrent vocabulary. However, our findings that speed of processing on adjective-noun trial does not explain unique variance when controlling for earlier vocabulary is in contradiction to this research.

Alternatively, other research has found that speed of linguistic processing is related to syntax acquisition (Fernald et al., 2006; Peter et al., 2019). When we consider late talking as failure to combine words together (a measure of syntax acquisition), it is possible that speed of processing would be a promising discriminator between late talkers and typically developing children. To explore this possibility, we ran an ROC curve analyses using speed of processing on adjective-noun trials using the failure to combine classification. The results of this analyses were poor; both sensitivity and specificity did not reach a satisfactory threshold. However, there were only five late talking children with reaction times available for adjective-noun trials. Therefore, more research is required, with more late talking children, to establish the role of speed of linguistic processing in language delay.

Our second set of analyses established the discriminatory ability of the six predictors that explained unique variance in the first set analyses, in classifying late talking and typically developing children. Using the 50 word cut off, most of the variables (earlier productive vocabulary, earlier receptive vocabulary, M3L and MWR) were successful in identifying late talking children with good levels of sensitivity. Only earlier gestures did not reach acceptable levels sensitivity when using the threshold of 70%. Using the failure to combine cut off, earlier gestures and NWR were successful in identifying late talking children with good levels of sensitivity. However, both earlier productive and receptive vocabulary did not reach acceptable levels of sensitivity.

When using the 50 word cut off, the findings from our discriminant function analyses show that NWR is the most important factor for discriminating between late talking and typically developing children. It should be noted that, in both analyses that include NWR, there were only 2 late talking children. In the one analysis without NWR, none of the late talking children were correctly identified.

Using the failure to combine cut off, we see that a combination of all of the variables can be used to discriminate between late talking and typically developing children. Again, NWR contributes the most the classification accuracy of this model. When using only non-experimental variables, the sensitivity does not reach our threshold for satisfactory. However, this analysis is substantially better than that using the 50 word cut off. Here, earlier receptive vocabulary contributes the most to the classification accuracy of this analysis. When looking at our best predictors (earlier gestures and NWR), this model correctly classified most children; all late talking children were correctly classified and 86.30% of typically developing children were correctly classified.

Child sex predicted unique variance in productive vocabulary at 25 months after controlling for earlier productive vocabulary scores. However, sex did not contribute a large proportion to the classification accuracy of late talking and typically developing children when using all of our measures or only non-experimental measures. A child’s sex does contribute to variance in early vocabulary development

(Horwitz et al., 2003; Nelson, 1973). Boys acquire vocabulary and syntax slower than girls (Huttenlocher et al., 1991; Ramer, 1976; Schachter, Shore, Hodapp, Chalfin, & Bundy, 1978). However, though there are more boys than girls in our low language sample, there is too much overlap in the number of boys in the two groups for child sex to be good at discriminating between late talking and typically developing children.

Language use at an earlier time point was successful at predicting both individual differences in productive vocabulary development and in classifying late talkers using the 50 word cut off. This is consistent with previous research showing language use as a strong predictor of individual differences (Henrichs et al., 2011; Westerlund et al., 2006) and late talking status (Klee et al., 1998; Reilly et al., 2007). It is unsurprising that a child's productive vocabulary at one time predicts variance in their productive vocabulary at a later point, and there is a large body of research in line with this finding (Henrichs et al., 2011; Rescorla, 2011). More interesting is the finding that earlier receptive vocabulary was a robust predictor of individual differences in later productive vocabulary and late talking using the 50 word cut off. The idea that a child's own lexicon facilitates the speed with which productive vocabulary grows has been previously proposed by Borovsky, Ellis, Evans and Elman, (2016) and by Jones & Rowland (2017). Borovsky et al. (2016) proposed that the larger a child's lexicon, the more concepts they understand and thus the easier they can recognise and learn new words. Jones & Rowland (2017) proposed that the more words a child knows, the more sublexical chunks they have stored in long term memory from which they can process and learn new material. Our finding that earlier receptive vocabulary was a robust predictor of individual differences in later productive vocabulary development even when we control for earlier productive vocabulary supports the idea that the words children know are having an impact on the rate of later language learning.

However, the finding that earlier language use was less successful in classifying late talkers according to the failure to combine criteria is surprising; a delay in early vocabulary development is one of the first indicators that a child may experience a language impairment and may cause parents to seek additional support for children's language development (Paul & Roth, 2011). However, combining words together is a measure of syntax ability. We know that there is a correlation between vocabulary development and syntax acquisition (Bates & Goodman, 1997; Rescorla, Dahlsgaard, & Roberts, 2000; Rescorla & Schwartz, 1990)., but it is possible that the effect of vocabulary on syntax is not big enough to classify late talking children on the basis of word combinations.

Of particular interest is the finding that M3L was a strong predictor of individual differences in productive vocabulary development and late talking status using the 50 word cut off. M3L is an early measure of syntax development, and there is a wealth of research showing there is a strong relationship between vocabulary and syntax at the beginning of syntax acquisition (Bates & Goodman, 1997; Rescorla et al., 2000; Rescorla & Schwartz, 1990). Since our vocabulary measures were not strong predictors of late talking status as defined by failure to combine (M3L), it is possible that the direction of this effect, at this age, is syntax impacting vocabulary development. When we consider that older children with persisting language impairments no longer exhibit a delay in vocabulary development but in syntax, and that 2 years is that age at which language delay begins to present, the direction of this relationship starts to make sense. If syntax is influencing vocabulary development, a delay in syntax acquisition could, for a time, impact vocabulary size.

Gesture use from 15-18 months predicted unique variance in productive vocabulary at 25 months and was successful at identifying late talkers who were not combining any words together. Interestingly, gesture use was not sensitive enough to correctly classify late talking children when using the 50 word cut off. It was also one of the weakest contributors to group classification in the discriminant function analyses using only non-experimental measures for the 50 word cut off classification. However, it contributed a larger proportion to group classification using the failure to combine cut off. It is possible that gesture scores are more successful than vocabulary scores at predicting variance in later vocabulary and late talking status because we find a floor effect in vocabulary between 15-18 months, and thus, gesture scores are actually a better measure of a child's actual communicative ability at this age.

Considering the individual differences and classification analyses together, NWR is most consistently robust predictor. We replicated findings that NWR predicts variance in vocabulary development and expanded these to show that it predicts unique variance in productive vocabulary development even when we control for earlier productive vocabulary and child sex. Of most interest, is the finding that NWR predicts unique variance in our full regression model and was the best discriminator for distinguishing between typically developing and late talking children. Jones, Gobet and Pine (2007) propose that NWR is a measure of vocabulary size and that performance on NWR tasks in a reflection of vocabulary size rather than phonological working memory. However, our finding that NWR predicted unique variance when controlling for earlier vocabulary and in our full model contradicts this conclusion. While NWR performance is undoubtedly affected by vocabulary size, our results suggest that something else – perhaps intrinsic capacity differences in phonological working memory (Gathercole & Baddeley, 1989) - also contributes. Similarly, the finding that NWR was the best measure for discriminating between late talking and typically developing children is in line with previous research that recommends NWR as a clinical marker for language impairment (Conti-Ramsden et al, 2001). However, one of the limitations of this finding is that most late talking children did not complete the NWR task. We can see from our discriminant function analyses that in the models that include NWR, only 2 late talking children had data when using the 50 word cut off, and only 5 when using the failure to combine cut off. Therefore, most of the late talking children do not actually complete the NWR task. Thus, in principle, NWR may be a promising predictor of late talking status, and have the potential to predict persisting late talking status. However, in practice it may be difficult to acquire reliable NWR results from children with low productive vocabulary scores.

3.5 Conclusion

We found that earlier receptive vocabulary and gestures, as well as concurrent M3L and NWR, predicted additional unique variance in productive vocabulary at 25 months after controlling for child sex and earlier productive vocabulary in a sample of late talking and typically developing children. Individually, earlier productive vocabulary, M3Land NWR correctly classified late talking and typically developing children using the 50 word cut off. Gestures and NWR correctly classified late talking and typically developing children using the failure to combine cut off. However, when combined, only NWR contributed to classification accuracy and without NWR these measures did not correctly distinguish between late talking and typically developing due to poor sensitivity. We conclude that individual measures are more successful at discriminating between these two groups of children.

4 Chapter 4. Follow-up study from the UK-CDI Project: Can risk factors, vocabulary skills and gesture scores in infancy predict later concern for language development?

Fit within the thesis

Chapters 2 and 3 examined if different factors could be used to identify children who are experiencing a delay in vocabulary acquisition. As discussed in chapter 1, it is difficult to discriminate between children whose vocabulary development does and does not catch up. Chapter 1 also details how the findings of previous research are mixed; the predictive power of early language development or risk factors, when examined alone, is often weak. This chapter examines the discriminatory ability of a combination early language skills (used in chapters 2 and 3), health and demographic risk factors, as well as early parental concern for language development (recorded at 15-18 months) to accurately distinguish between children for whom there is and is not concern for their language development in later childhood (recorded at 4-6 years).

The study reported in this chapter found that, while these factors could successfully identify children for whom there was no concern for their language development, earlier language skills, health and demographic risk factors, and early parental concern for language development did not successfully identify children for whom there was overall concern for their language development, either individually, or when combined together. This was due to a lack of sensitivity in each model.

This study was designed by Lana Jago and Caroline Rowland. This study was a follow-up of the UK-CDI project. Data for the UK-CDI Project was collected across the UK between 2013-2015. Lana Jago collected the data for the follow-up project, analysed the data and wrote the first draft of the paper. Caroline Rowland gave comments and revisions on the paper. This chapter is in preparation for submission to a peer reviewed journal (Jago, Pine & Rowland, in prep.).

Abstract

This study followed up families who took part in the UK-CDI Project to investigate if a combination of health and demographic risk factors, as well as vocabulary and gesture scores recorded between 15-18 months, could be used to identify children who have had concern expressed for their language development at 4-6 years. A questionnaire was used to measure concern for children's language development and children were identified as having concern for their language development if their parents answered in the affirmative to at least one of five questions: parental concern about language development, diagnosis of developmental disability, diagnosis of language disorder, hearing or visual impairments, or identification of early language delay at the Healthy Child Programme's 2 Year Review. Discriminant function analyses revealed that while these factors could successfully identify children without concern for their language development, they were not sensitive enough to identify children for whom there was concern for their language development.

4.1 Introduction

Delays in language development and impairment over time are associated with exposure to more adverse health and demographic risk factors (Campbell et al., 2003; Paul, 2000; Reilly et al., 2010). In addition, young children with early delays in language acquisition are at an increased risk for developing persisting language impairments (Rice, Taylor, & Zubrick, 2008).

It is widely accepted that risk factors impact language development, with more risk factors being associated with poorer language development (Law, McBean, & Rush, 2011). As a result, policy-makers and lobby groups often recommend that the language development of children growing up with greater disadvantage is monitored through their first years and that they receive targeted support aimed at improving their communication environments (All Party Parliamentary Group on Speech and Language Difficulties, 2013; Save the Children, 2015).

However, while there is a wealth of research examining the role of environmental risk factors in predicting later language outcomes (Reilly et al., 2007; Reilly et al., 2010), the results of these studies have not been conclusive. Some research has found slower language acquisition and increased risk for persisting language impairment is associated with early identified health and demographic risk factors (Campbell et al., 2003), but other research has failed to find this relationship (Reilly et al., 2010). This means we are currently unable to successfully predict later language outcomes from infancy using exposure to risk factors as a predictor (Bishop et al., 2012; Duff, Nation, Plunkett, & Bishop, 2015; Roos & Ellis Weismer, 2008).

In the remainder of this introduction, we outline some of the contradictory research around the role of risk factors, early concern for language development and earlier language skills in predicting later language impairment. Following this, we describe three reasons why this research is inconclusive, before describing the current study, the aim of which is to provide additional evidence, from a large representative sample of children, about the role of early risk factors in predicting later language delay. Note that in our study, we use professional or parental concern for language development when the child is aged 4 to 6 years as our index of later language delay, rather than a diagnosis of language disorder. The reasons why we do this are explained below.

4.1.1 Language proficiency

It is widely accepted that a child's own language ability at time 1 predicts their language ability at time 2, at least in children within the typical range (Henrichs et al., 2011; Rescorla, 2011; Westerlund, Berglund, & Eriksson, 2006). However, the role of language proficiency in predicting language impairment is less clear cut. Many children who experience an early delay in language acquisition are later diagnosed with a developmental language disorder. However, since many other children do catch up, language ability at an early age is not a robust predictor of later language delay. For example, Westerlund et al (2006) found that a measure of vocabulary size at 18 months was not sensitive enough to identify children with language impairments at 3 years.

However, it is possible that combining a number of early language measures (e.g. combining productive vocabulary with measures of receptive vocabulary, gestures, risk factors and early concern for language impairment) will improve the discrimination between children for whom there is, and is not, concern for their language development. Measuring only early productive vocabulary, especially in children who experience a delay in vocabulary acquisition, may result in a floor effect. Therefore, receptive vocabulary may be a more accurate measure of children's communicative abilities when they are young and, indeed, previous research has shown that receptive vocabulary predicts later productive vocabulary development (Duff et al., 2015). However, the role of receptive vocabulary in predicting later language impairment is relatively unexplored. We hypothesise that receptive vocabulary is likely to be good at discriminating between children who do and do not have delayed language later on, as indexed by parental or professional concern for their language development, because it will capture children's early communication skills.

Measures of early gesture use may, similarly, be a robust way to measure children's early communicative skills, especially since we know they are related to later language development and language impairment (Hsu & Iyer, 2016; Colonnesi, Stams, Koster, & Noom, 2010 ;Rowe & Goldin-Meadow, 2009). Hsu and Iyer (2016) found children at risk for language impairment at 3 years produced fewer gestures at 15 months. However, this effect disappeared at 4 and a half years, which suggests that the relationship between early gesture use and later language development may hold only for a few years, with other skills becoming more important as the child ages That said, while gestures alone many not be strong enough to predict later language impairment over time, the combination of gesture scores with vocabulary scores and with other risks factors may improve our ability to discriminate between children for whom there is and is not concern for their language development in later childhood.

4.1.2 Health risk factors

A variety of health factors have been shown to be associated with language development over time (child sex, prematurity, low birth weight, ear infections, family history of speech or language impairments, and developmental disabilities; Barre, Morgan, Doyle, & Anderson, 2011; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991). For example, Reilley et al. (2010) found that multiple health related risk factors (child sex, low birth weight and family history of speech of language difficulties) predicted variance in language skills at age 4 years. Similarly, Jansson-Verkasalo et al. (2004) showed that being born premature was associated with poorer performance on measures of comprehension at 2 years.

However, some research investigating the role of such risk factors in predicting later language impairment often finds weak effects (Harrison & McLeod, 2010). For example, Harrison and McLeod (2010) found that while risk factors correctly distinguished between children who were and were not attending speech-language pathology services, they only correctly classified 2.6% of children who were attending these services. Therefore, risk factors were not sensitive enough to identify children with language impairment. Here again, however, it is possible that we can improve classificatory accuracy by combining risk factors with other measures, such as early vocabulary and gesture scores.

4.1.3 Demographic risk factors

Two demographic factors which have been shown to predict language development over time are maternal education and household income. It is likely that these factors impact language development together because they interact; rate of household income can be dependent on maternal education and they both, together, determine socioeconomic status (SES). Previous research has shown that maternal education impacts language development via the mother's own language skills; parents with higher educational attainment have more advanced literacy skills and are therefore more likely to produce rich linguistic input from which children learn (Rowe, 2018; Rowe, Pan, & Ayoub, 2009). However, other research has found that the effect of maternal education on predicting later language is weak (Harrison & McLeod, 2010).

Socioeconomic status, as established via household income, has been shown to impact language development via quantity of parental input; parents from lower SES backgrounds do not have the same amount of opportunity to provide the wealth of input children from higher SES background receive. For example, parents living in more chaotic and crowded conditions use less varied language with their children (Evans, Maxwell, & Hart, 1999; Rowe, 2018). Reilly et al. (2010) found that maternal education and socioeconomic disadvantage (measured using household income) predicted both performance on language assessments and language impairment at age 4 years. However, Harrison and McLeod (2010) found family income did not predict whether or not children were attending speech-language pathology services. It is, nonetheless, possible that SES factors, though unreliable on their own, when combined with health risk factors, early language skills and early concern for language development will be strong predictors of language outcomes.

4.1.4 Early concern for language development

Parental concern for language development early in childhood has also been shown to support the identification of language impairment (Glascoe, 1991). When combined with clinical observations, parental evaluation of their children's language development increases the accuracy with which paediatricians can detect developmental complications (Glascoe, Altemeier, & MacLean, 1989; Glascoe & Dworkin, 1995). In addition, a delay in vocabulary development is one of the reasons that parents first seek support for their children's vocabulary development (Rescorla, 2011). Therefore, parental concern could be used in combination with other risk factors to support the detection of delays in language development before language impairment is identified by a clinician or therapist.

In sum, the research outlined above shows that the role of risk factors in predicting language outcomes and impairment is somewhat contradictory. There are a few possibilities which can explain why these results are unreliable. First, it is possible that the effect of these risk factors in language acquisition is quite weak overall, and therefore, only very high-powered studies with large sample sizes are successful in finding relationships. Where effects are found, these very often are weak (Carroll & Breadmore, 2017; Stolt, Haataja, Lapinleimu, & Lehtonen., 2009), and studies with more power are often more successful detecting relationships between risk factors and language acquisition (Barre et al, 2011; Winksel, 2006; Kennedy et al., 2006; Van Noort-Van Der Spek, Franken, & Weisglas-Kuperus, 2012).

Second, it is possible that risk factors are only strong predictors of later language outcomes at the group level but fail to predict on the individual level and are, in turn, weak predictors of language disorders. It is not inevitable that predictors of individual differences are the same as predictors of language disorders. Some measures may discriminate well between faster learners, or between learners in general, but have very little predictive power when identifying children whose productive vocabulary is of lower ability. For example, boys tend to be slower than girls on average but the differences are so small, and the overlap in standard deviation so large, that sex has very little discriminatory power for identifying children whose productive vocabulary is of lower ability. In support of this view, Reilly et al. (2010) found that nine risk factors were successful in predicting individual differences in language scores at 4 years, but only three of them allowed the authors to discriminate between children with and without expressive specific language impairment (SLI; now Developmental Language Disorder (DLD)).

Third, it might be possible that risk factors, when examined in isolation, are not strong enough to predict language impairment (e.g. Harrison & McLeod, 2010) but gain substantial predictive power when combined. However, very little research has investigated how earlier language skills and multiple risk factors might interact to increase the risk of developing persistent language impairments. Using multiple risk factors, combined with measures of earlier language skills, could increase our chances of successfully predicting language impairment. Therefore, it is important to investigate risk factors and early language skills together. If we can show that a combination of risk factors and language skills identified early can predict later language impairment, we can use these factors as a starting point for informing early intervention based on risk, before language impairment reveals itself.

4.1.5 The current study

This study followed up a large unique cohort of children who participated in the UK-CDI project in later infancy (15-18 months). We investigated whether a combination of environmental risk factors, early language skills, and whether parents had expressed any worries about their child's language in late infancy, could distinguish between children whose language development in later childhood (between 4 and 6 years of age) was of concern. 'Concern' about language development was defined as parents answering in the affirmative to one or more of the following questions: a) having being concerned about their child's language development in the past, b) their child having had a diagnosis of developmental delay, c) a diagnosis of language disorder, or d) a visual or hearing language delay at the Healthy Child Programme's Two Year Review (a UK-wide review of development that occurs between 24 and 30 months of age, performed by a health professional). We chose this broad category of 'concern', rather than a diagnosis of language disorder as our outcome variable, as many children will not yet have a diagnosis at the age at which we tested them (4-6 years). In addition, as stated above, parental concern for language development has been shown to benefit clinical evaluations of language impairment.

The cohort were a subset of the parents who had taken part in the UK-CDI Project when their children were between 15 and 18 months (time 1). The UK-CDI Project collected parental report data from across the United Kingdom to establish norms for productive and receptive vocabulary, and gestures for children from 8-18 months. Parents completed a vocabulary and gesture questionnaire (a Communicative Development Inventory, or CDI), and a family questionnaire containing questions about child health, familial risk of language and literacy disorders, and demographic characteristics. The original cohort were representative of the UK population as a whole in terms of a range of demographic factors (e.g. socio-economic status, sex, region, nation, marital status etc). For the present study, those parents who had agreed to be re-contacted were asked to complete a follow-up questionnaire when their children were between 4 and 6 years (time 2). The follow-up questionnaire determined whether anyone had expressed concern about the children's language development at time 2, with concern defined according to the five categories outline above.

The study had three aims. First, we wanted to determine whether overall concern for language development at time 2 was predicted by parental concern for their child's language development between 15-18 months (time 1). We anticipated that parents' concerns about their children's language development in late infancy would predict later overall concern for children's later language development. Second, we investigated whether a combination of risk factors and early language and gesture scores at time 1 predicted an increased likelihood of there being concern expressed for children's language development at time 2. We hypothesised that a combination of risk factors and earlier language skills would be a better predictor of later overall concern for language development than any individual risk factors or language skill in isolation. Third, we investigated if risk factors and early language skills could predict whether or not children's language development caught up after concern for their language development had been identified. We anticipated that children with higher vocabulary and gesture scores at time 1, as well as those who experienced fewer risk factors at time 1, would be more likely to catch up compared to children experiencing more risk factors and having lower vocabulary and gesture scores.

4.2 Methods

4.2.1 Participants

The original data were collected as part of the UK-CDI Project, a large-scale project aimed at creating norms for productive vocabulary, receptive vocabulary and gestures for children in the UK from 8-18 months (Alcock, Meints, Rowland, Brelsford, Christopher, & Just, 2020). This project collected vocabulary and gesture data from 1210 children aged 8-18 months across the UK. Parents also provided information about their children's health and family background. The original sample was designed to be representative of the population of UK children of this age in terms of a number of demographic factors including sex, region, SES, and race

We contacted all of the families from the original data collection phase who had agreed to be contacted for further studies, had provided contact details, and whose children were between 15-18 months when they completed the UK-CDI (n=370; 78 of whom had productive vocabulary scores in the bottom 25th percentile). We received 147(40%) responses; 223 families did not reply (60%).One family was excluded because their response was incomplete, so the final sample size was 146 (see table 4.4 for the demographic information for all families contacted, split by those who did and did not reply to the follow-up questionnaire). Out of the final sample of 146 children that contributed data for the follow-up project, 24 had a family history of language delay or dyslexia. Around half of the cohort were girls (n= 70). Twelve families had low maternal education and 34 had low family income. Details on how these cut-offs were established is in section 4.2.5.2, below. Of the 223 families who did not reply, 113 were of the children were girls. Sixty-three had low maternal education and 104 had low family income.

The mean age of children 146 included in the follow-up project was 16 months 25 days (15 months, 3 days -18 months, 28 days) during the UK-CDI Project (time 1), and 5 years and 3 months (4 years, 3 months - 6 years, 4 months) at follow-up (time 2). All children were monolingual English learners. Details on the number of children who had ear infections is given in table 4.1.

4.2.2 Overall design

We used a questionnaire design to follow up children who previously took part in the UK-CDI Project, and had agreed to be contacted for future studies. This study was granted ethical approval by The University of Liverpool's Research Ethics Subcommittee for Non-Invasive Procedures for the study Language Development in Late Talkers (Institute Review Board protocol number: RETH000764).

We ran four sets of analyses. In the first analysis, we tested whether the families who replied to the follow-up questionnaire were representative of the entire UK-CDI Project sample. We ran preliminary Chi^2 analyses to assess the rate of risk factors reported by families in both groups. In the second analysis, we used logistic regression to test if parental concern at time 1 increased the likelihood of overall concern at time 2. For this analysis the independent variable was parental concern at time 1 and the dependent variable was overall concern at time 2. In the third set of analyses, we ran five discriminant function analyses to establish if different combinations of risk factors, together with vocabulary and gesture scores recorded at time 1, were associated with increased likelihood for concern at time 2. The independent variables for these analyses were language group at time 1, health risk factors (child sex, prematurity, low birth weight, ear infections at time 1, familial risk for speech or language impairment, ear infection lasting more than three months, developmental disability), demographic risk factors (maternal education, household income), and concern expressed at time 1 (hearing or communication concerns at time 1), and finally, all of these variables combined. The dependent variable for these analyses was overall concern for language development established at time 2. We then ran these five analyses again using a more stringent criteria of concern that included only the questions that asked about a diagnosis of a disability: a diagnosis of a developmental disability, a diagnosis of a language disorder, or a visual or hearing impairment.

In the fourth set of analyses, we tested whether our risk factors, when combined, could predict 'catch up ability' (i.e. could distinguish between children whose language had been of concern at some point in their development but whose difficulties resolved by time 2, and those whose language was still of concern). This included only the subset of children whose parents answered yes to the question "Have you ever worried that your child's speech was delayed compared to other children the same age?" at time 2. For this analysis, the independent variables were the risk factors listed above, as well as vocabulary and gesture scores at time 1. The dependent variable was the answers to the language catch-up question at time 2 (Did your child's speech eventually catch up with up to that of other children the same age?).

4.2.3 Sampling and data collection procedures

The first data collection (time 1) took place as part of the UK CDI Project from 2013 to 2015. The data for the follow-up project (time 2) was collected between 2017 and 2018. We used the database from the UK-CDI Project to follow-up families who had consented to be re-contacted for future research. Parents were given a £5 shopping voucher for completing the follow-up questionnaire.

4.2.4 Measures and procedure

4.2.4.1 The UK-CDI

Communicative Development Inventories (CDI) are parent report checklists of words and gestures/sentences. Parents complete these questionnaires by indicating if their child uses or understands the words and gestures listed. The UK-CDI Words and Gestures is standardised for the UK population for vocabulary and gesture scores in children aged 8-18 months and has good validity and reliability (total possible scores are 396 for vocabulary and 57 for gesture).

4.2.4.2 The Family Questionnaire

The family questionnaire asks a range of questions about a child's health and family background. This questionnaire was designed for the UK-CDI Project (for details of construction, see Alcock et al, 2020). This questionnaire was used to collect information about demographic and health risk factors, including prematurity, birth weight, family history of language delay or dyslexia, and SES. Questions of interest included "1. At what week of pregnancy was your child born?", "2. How much did your child weigh at birth?", "3. Has your child had an ear infection/glue ear for longer than 3 months, 4 to 6 ear infections within a 6 month period, or another identified hearing problem (e.g. at newborn hearing screening)?", "4. Is there anyone in the child's immediate family (brothers/sisters/parents only) with a speech/language difficulty?", "5. Does your child have a developmental disability (e.g. Cerebral Palsy, ASD, Fragile X syndrome, Muscular dystrophy, Di George syndrome, Down's syndrome, Williams syndrome)?", "6. Does your child have a hearing or visual impairment?", "7. Have you or anyone else had any concerns about your child's hearing or communication?", "17. Mum's highest education is a. No formal qualifications, b. GCSE/O Level/NVQ Level 1 or2/ similar, c. A Level/NVQ Level 3/similar, d. University degree/HND/HNC/NVQ Level 4 or 5/similar, d. Postgraduate/similar e.g.(PGCE, PhD, MA etc.)", "27. What is the overall household income (before tax) per year in your child's main home? a. £0-£14000, b. £14,001-£24,000, c. £24,001-£42,000, d. £42,001 or more".

4.2.4.3 Follow-up questionnaire

The follow-up questionnaire was used to investigate the language outcomes of the children who took part in the UK-CDI project. The key questions for this study are those that asked about parental concern for language development, details of those concern (if any), whether the Healthy Child Programme's 2 Year Review identified a delay in language development, whether the children had been diagnosed with a developmental disability or language disorder, and whether the children had a visual or hearing impairment. The Healthy Child Programme 2 Year Review (Department of Health, 2009) is part of the *Healthy Child Programme: Pregnancy and the first five years of life*, which is run in England and Wales. This review is designed to optimise child development by reviewing all children in England and Wales between 2 years and 2 years, 6 months.

At time 2, parents who had provided an email address at time 1 were sent an email containing a link to complete the questionnaire online. For parents who only provided a home address, a paper copy of the questionnaire was sent out with a prepaid return envelope included. See the supplementary materials at https://osf.io/gvz3x/ for a copy of this questionnaire.

4.2.5 Data coding

4.2.5.1 Concern scores at time 2

Five questions on the follow up questionnaire at time 2 were used to create 'concern' scores:

1. Parental concern for language development: answering yes to “have you ever worried that your child’s speech was delayed compared to other children the same age?” = score of 1
2. Developmental disability: answering yes to “does your child have a developmental disability?” = score of 1
3. Diagnosis of language disorder: answering yes to “has your child been diagnosed with any of the following language disorders?” = score of 1
4. Hearing or visual impairment: answering yes to “does your child have a hearing or visual impairment?” = score of 1
5. Identification by the Healthy Child Programme’s 2 Year Review (the Two Year Check): answering yes to “did this programme identify any delays with your child’s speech, language or communication abilities?” = score of 1

The parents’ answers to these questions were used to calculate two scores for each child:

6. Overall concern: Children whose parents answered yes to any of the question above were given a concern score of 1. Those whose parents answered no to all of the questions above were given a concern score of 0.
7. Identified disability: Children whose parents answered yes to any one of questions 2, 3 or 4 above were given a score of 1. Children whose parents answered no to questions 2, 3 and 4 were given a score of 0.

Forty-nine children (33.56%) were identified as having an overall concern score of 1 at time 2. Twenty children (13.70%) fit the criteria for having an identified disability (developmental disability =10, language disorder =1 and/or visual or hearing impairment= 12). Note that DLD is estimated to affect approximately 7.58% of the population (Norbury et al., 2016).

Parents who answered yes to the question “have you ever worried that your child’s speech was delayed compared to other children the same age?” were also asked a catch-up question (“Did your child's speech eventually catch up with that of other children the same age?”). Of the 37 parents who expressed concern for their children’s language development, 26 reported that their children’s language had caught up with children the same age, and 11 reported that it had not caught up.

4.2.5.2 Risk factor scores at time 1

Information about 10 risk factors were collected at time 1:

1. Concern (at time 1): answering yes to “have you or anyone else had any concerns about your child’s hearing or communication?” = score of 1

Physical/health factors:

2. Child sex: operationalised as being male= score of 1

3. Prematurity: operationalised as being born before week 36 = score of 1
4. Low birth weight: operationalised as weighing less than 5lb 8oz when born = score of 1
5. Ear infection: operationalised as answering yes to the question “has your child had an ear infection/ glue ear for longer than 3 months, 4 to 6 ear infections within a 6 month period, or another identified hearing problem?” = score of 1
6. Familial risk of language/literacy disorder: answering yes to “is there anyone in the immediate family with speech/language difficulty or dyslexia?” = score of 1
7. Developmental disability: answering yes to “does your child have a developmental disability?” = score of 1
8. Hearing or visual impairment: answering yes to “does your child have a hearing or visual impairment?” = score of 1

Demographic factors:

9. Maternal education: selecting “no formal qualifications” or “GCSE/O level/NVQ level 1 or 2” = score of 1
10. Household income: selecting “£0-£14,000” or “£14,000-£24,000” = score of 1

The cut off scores for maternal education and household income used above were designed to determine low SES status. For household income, families with income of around £22,800 per year are considered to have low income (Department for Work and Pensions, 2019). Low maternal education was established as having no formal qualifications or GCSE/O level/ NBQ level 1 or 2. Previous research has shown that children of mothers with less than 12 years of education are at increased risk for persisting language impairment (Stanton-Chapman, Chapman, Bainbridge, & Scott, 2002).

4.2.6 Language measures at time 1

4.2.6.1 Division by quartiles: Group membership

Because we wanted to be able to identify if being in the bottom 25th percentile for vocabulary/gesture at time 1 would predict concern at time 2, we divided the children into four groups based on vocabulary and gesture scores between 15-18 months using the UK-CDI norms. The UK-CDI norms were created using the entire UK-CDI Project sample and provide percentile cut-offs for productive and receptive vocabulary, and gestures for each month. Children were split into four groups based on percentiles: 0-25th, 25th-50th, 50th-75th, and above

75th percentiles. Each child was placed into one of the 4 groups separately for productive vocabulary, for receptive vocabulary and for gesture use.

4.2.6.2 Language Z-Scores

Standardised scores for productive vocabulary, receptive vocabulary and gestures were calculated using the data from the UK-CDI Project data using the mean and standard deviation (SD) of gesture, productive and receptive vocabulary scores for the entire standardisation sample at all four ages (15, 16, 17 and 18 months). These are age-adjusted scores which control for the fact that the children were different ages at time 1 (between 15 and 18 months).

4.2.7 Analysis strategy

All analyses were conducted using one-tailed tests, as all hypotheses are unidirectional hypotheses. All outliers were included, unless it was determined that the data point was due to experimenter or participant error. Chi^2 analyses were performed in R (R Core Team, 2017, R version 3.4.1) using R Studio (Version 1.0.153) using the CrossTable function as part of the gmodels package (Warnes, Bolker, Lumley & Johnson, 2018). Logistic regressions were performed in R using the glm function as part of the pscl package (Jackman, 2010). Discriminant function analyses were run in SPSS Statistics 24.

4.3 Results

The aim of this study was to (1) assess if parental concern at time 1 was accurate enough to predict overall concern at time 2, (2) investigate whether a combination of risk factors and earlier vocabulary and gesture scores can predict later concern for language development and (3) investigate if risk factors and earlier vocabulary and gesture scores can distinguish between children whose language does or does not catch up after parents expressed concern for their language development.

4.3.1 Descriptive statistics

Descriptive statistics of all the children whose parents provided data for the follow-up questionnaire can be seen in table 4.1. Table 4.2 details the number of children in each quartile group for productive vocabulary, receptive vocabulary and gestures from 15-18 months. Table 4.3 details the number of children whose language did or did not catch up when their parents reported that they had concern for their language development at time 2.

Table 4.1:

Number of children with and without each risk factor

	Girls			Boys		
Risk factor	*n* with the risk factor	*n* without the risk factor	Missing data	*n* with the risk factor	*n* without the risk factor	Missing data
Overall concern at time 2 (answering "yes" to any of the 5 questions which denote concern)	16	54	0	33	43	0
Identified disability (answering "yes" to the three questions on diagnosis of developmental disability, diagnosis of language impairment, having a visual or hearing impairment at time 2)	4	66	0	16	60	0
Concern expressed by parent at time 2	10	60	0	27	49	0
Concern expressed at Two Year Review time 2	6	56	8	16	45	15
Diagnosis of developmental disability time 2	3	67	0	7	69	0
Visual or Hearing impairment time 2	1	69	0	11	65	0
Diagnosis of language disorder time 2	0	70	0	1	75	0
Health problems at time 1 (total)						
Prematurity time 1	5	65	0	8	68	0
Low birth weight time 1	6	64	0	5	71	0
Ear infection at time 1	1	69	0	2	74	0
Familial risk (someone in family) time 1	12	58	0	12	63	1
Developmental disability time 1	0	69	1	0	76	0
Visual or hearing impairment time 1	1	69	0	1	73	2
Language concerns at time 1						
Hearing or communication concerns at time 1	2	68	0	8	68	0
Demographic factors at time 1 (total)						
Maternal education time 1	9	61	0	3	73	0
Household income time 1	18	52	0	16	60	0

Table 4.2:

Number of children in each quartile for vocabulary and gestures at time 1 (group membership)

Variable	0-25th	25th-50th	50th-75th	75th-100th
Productive Vocabulary	39	30	39	38
Receptive Vocabulary	33	32	32	49
Gestures	30	36	43	37

Table 4.3:

Number of children whose parents reported that their language has or has not caught up by time 2

Caught up	Not caught up
26	11

To check if any of the variables were highly collinear we ran Chi^2 analyses between each variable. For any two variables that yielded significant Chi^2 scores, we followed this up with a Cramer's V post-test to establish the collinearity. Cramer's V provides an effect size where values vary between 0 and 1, with 0 indicating no collinearity and 1 indicating high collinearity. The Chi^2 analyses revealed eight of the predictor variables were significantly associated (prematurity and low birth weight; low birth weight and family history of language delay or dyslexia; ear infection at time 1 and visual or hearing impairment at time 1; family history of language delay or dyslexia and a visual or hearing impairment at time 1; family history of language delay or dyslexia and maternal education; visual or hearing impairment at time 1 and maternal education; maternal education and household income; productive vocabulary group and receptive vocabulary group; productive vocabulary and gesture group; receptive vocabulary and gesture group). However, none of these were variables were highly collinear (all Cramer's V values below 0.70). Collinearity between developmental disability at time 1 and all other variables could not be established because no parents reported developmental disability at time 1.

4.3.2 Preliminary analysis: Demographics of families who did and did not reply to the follow-up questionnaire

We ran Chi^2 analyses to investigate if there were any differences in the risk factors listed in table 4.4, recorded at time 1, between families who did and did not respond to the follow-up questionnaire. There were no significant differences between the groups in terms of prematurity, birth weight, incidence of ear infections, family history of language delay or dyslexia, developmental disability, visual or hearing impairment and hearing or communication concerns. However, there were significant differences between response groups in maternal education (χ^2 (4)=23.25, *p*=.0001) and household income (χ^2 (3)=22.03, *p*=.0002). Families with higher maternal education and household income were more likely to respond to the follow-up questionnaire. The results for these analyses, along with percentages of children in each response group who had each risk factor at time 1, can be seen in table 4.4.

Table 4.4:

Number and percentage of children in each risk factor category whose parents did and did not reply. Results are split by sex . The χ^2 analysis is a comparison of all families who did and did not reply, not split by sex

	Girls		Boys				
Risk factor	Families who replied (%)	Families who did not reply (%)	Families who replied (%)	Families who did not reply (%)	χ^2	*df*	*p*
Health problems							
Prematurity time 1	5 (7%)	6 (5%)	8 (11%)	11 (10%)	1.86	2	.39
Low birth weight time 1	6 (9%)	8 (7%)	5 (7%)	5 (5%)	0.77	2	.68
Ear infection at time 1	1 (1%)	1 (1%)	2 (3%)	7 (6%)	0.73	1	.39
Familial risk time 1	12 (17%)	17 (15%)	12 (16%)	19 (17%)	0.01	1	.92
Developmental disability time 1	5 (7%)	6 (5%)	8 (11%)	11 (10%)	0.67	1	.41
Visual or hearing impairment	6 (9%)	8 (7%)	5 (7%)	5 (5%)	0.71	1	.40
Language concerns at time 1							
Hearing or communication concerns at time 1	1 (1%)	1 (1%)	1 (1%)	5 (5%)	0.04	1	.85
Demographic factors							
Maternal education time 1	2 (3%)	4 (4%)	8 (11%)	10 (9%)	23.25	4	.0001
Household income time 1	2 (3%)	4 (4%)	8 (11%)	10 (9%)	22.03	3	.0002

4.3.3 Main analysis 1: Strength of parental concern over time

We ran a logistic regression analysis to investigate if parental concern at time 1 predicted overall concern at time 2. Parental concern at time 1 did not increase the likelihood of overall concern at time 2, b= 1.18, SE=0.67, p=.08, but note that, only 10 parents expressed concern at time 1.

4.3.4 Main analysis 2: Predicting overall concern for language at time 2 from risk factors, language skills and gesture scores recorded at time 1

Next, we ran discriminant function analyses to assess the discriminatory ability of the risk factors and language group at time 1 in correctly classifying children into two groups. Discriminant function analysis is a statistical technique used to determine how well predictor variables discriminate between two or more naturally occurring groups. Here we used it to determine which different combinations of risk factors gave us the best classification accuracy of children into our two outcome groups: children for whom concern about language has been expressed (1) and children for whom concern about language has not been expressed (0) by time 2. Discriminant function analysis yields an overall accuracy figure (how well the model performs at discrimination overall), and sensitivity and specificity values. The sensitivity value measures the ability of the model to correctly classify children for whom there is concern for language development (true positives). Specificity measures the ability of the model to correctly classify children for whom there has been no concern expressed about their language development (true negatives). Sensitivity and specificity rates between 70-80% are deemed acceptable for diagnostic assessments (e.g. for autistic spectrum disorder; Bright Futures Steering Committee & Medical Home Initiatives for Children With Special Needs Project Advisory Committee, 2006). Therefore, we consider accuracy, sensitivity and specificity and values above 70% to be adequate in the present analyses. Discriminant function analyses also provides standardised canonical coefficients for each variable. These coefficients allow us to compare the weighted importance of each variable in predicting group membership.

We ran five analyses (see table 4.5 for the overall results, and table 4.7 for the standardised canonical coefficients for each variable in each model, which indicate the weighted importance of each variable in predicting group membership).

The first discriminant function analysis included language group at time 1 (quartile groups for productive vocabulary, receptive vocabulary and gestures at time 1) predicting group membership (Overall Concern, No Overall Concern). This model correctly classified children into their groups with an accuracy of 67.10%, r=0.35 χ^2=18.89, df= 3, p<.001, and had good specificity (80.40%) meaning that it did well in classifying children in the No Overall Concern group (i.e. those for whom no concern had been expressed at time 2). However, the sensitivity was poor at 40.80%, so the model did not do well in identifying the children in the Overall Concern group at time 2. We also reran these analyses predicting concern group membership using productive vocabulary, receptive vocabulary and gesture z-scores, rather than language quartile groups. The results from these analyses were essentially the same.

The second discriminant function analysis tested the effect of health risk factors at time 1: child sex, prematurity, low birth weight, ear infections, familial risk for speech or language impairment, developmental disability, and visual or hearing impairments. This model correctly classified children into their groups with an accuracy of 70.80%, r=.33 χ^2=15.40, df= 6, p=.02. Again, although the specificity of the complete model was excellent, 97.90% meaning it correctly classified almost all children in the No Concern Group, it did not do well in terms of correctly classifying children in the Overall Concern group (sensitivity = 14.90%).

The third discriminant function analysis tested the effect of demographic factors predicting group membership: maternal education and household income. This model correctly classified children into their groups with an accuracy of 71.20%, r=0.25 χ^2=9.48, df= 2, p=.009. Again, however, specificity was good (89.70%,) but sensitivity was poor (34.70%).

The fourth discriminant function analysis included only parental concern for language development at time 1 predicting group membership. This model failed to correctly classify children into their groups, with an accuracy of only 67.80%, r=.15 χ^2=3.35, df=1, p=.07. Again, the sensitivity of this model was poor, 12.20%, therefore it did not do well in terms of correctly classifying children with overall concern for their language development at time 2. The specificity of the model, however, was excellent, 95.90%, it correctly classified most children without overall concern for their language development at time 2.

The fifth discriminant function analysis included all risk factors and language group at time 1, to determine if, together, these variables can be used to predict group membership. The included risk factors were health factors (child sex, prematurity, low birth weight, ear infections at time 1, familial risk for speech or language impairment, developmental disability, visual or hearing impairment), demographic factors (maternal education, household income), concern expressed at time 1 (hearing or communication concerns at time 1), and language groups at time 1 (quartile groups for productive and receptive vocabulary and for gestures). This model correctly classified children into these two groups with an accuracy rate of 75.70%, r=0.50, χ^2=39.26, df= 12, p<.001. Again, however, as with all the previous models, though specificity was good (91.80%), sensitivity was poor (42.60%). In sum, all of the models had low sensitivity, and were thus unable to classify children into the Overall Concern group with reliable levels of accuracy.

However, our Overall Concern measure identified a larger number of children whose language raised concern (33.56%) than we might expect, given the prevalence of language disorder in the population (e.g. DLD is estimated to affect approximately 7.58% of the population, Norbury et al., 2016). Thus, we ran these five analyses again with a stricter criteria, testing the models ability to classify those children who had/had not been identified with a disability that might affect language (being diagnosed with a developmental disability, being diagnosed with a language disorder and/or having had a hearing or visual impairment). Children were split into two groups: children for whom a problem which may impact their language development has been identified (Identified Disability = 1) and children for whom there has been no disability identified by time 2 (Identified Disability = 0). Here, the sensitivity value measures the ability of the model to correctly classify children for whom there is an identified disability (true positives). Specificity measures the ability of the model to correctly classify children for whom there is not an identified disability (true negatives). The results from these five analyses can be seen in tables 4.6 and 4.7. As with the previous five analyses, the sensitivity of these models was very poor; they were unable to accurately classify children with an Identified Disability.

4.3.5 Main analysis 3: Predicting catch-up in language development from risk factors recorded at time 1

For this analysis, we ran a discriminant function analysis to assess the discriminatory ability of the risk factors and language group in correctly classifying children whose language did (0) and did not (1) catch up by time 2. A total of 37 parents expressed concern for their children's language development at time 2. Of these 37, 26 reported that their children's language had caught up with children the same age, and 11 reported that it had not caught up.

This model did not correctly classify children into their groups. While the accuracy of this model was good, 82.90%, it did not reach significance, r=0.57 χ^2=10.66, df= 11, p=.47. This result is reflected in the sensitivity of the model. The sensitivity of this model was poor, 54.50% meaning it did not do well at classifying children whose language did not catch up. The specificity of the model was good, 95.80%, meaning it did well in terms of correctly classifying children whose language did catch up. See table 4.8 for the results of the discriminant function analysis and table 4.9 for the standardised canonical coefficients for each variable.

Table 4.5:

Results from the discriminant function analyses distinguishing between children with and without overall concern for their language development at time 2.

Variable	*r*	χ^2	n	*df*	*p*	Accuracy	Sensitivity	Specificity
Language group at time 1 (quartile groups for vocabulary and gesture scores at time 1)	0.35	18.89	146	3	<.001	67.10%	40.80%	80.40%
Health factors (sex, prematurity, low birth weight, ear infection, visual or hearing impairment, family history, developmental disability)	0.33	15.40	143	6	.02	70.80%	14.90%	97.90%
Demographic factors (maternal education, family income)	0.25	9.48	146	2	.009	71.20%	34.70%	89.70%
Concern at time 1	0.15	3.35	146	1	.07	67.80%	12.20%	95.90%
All variables	0.50	39.26	143	12	<.001	75.70%	42.60%	91.80%

Table 4.6:

Results from the discriminant function analyses distinguishing between children with and without an Identified Disability at time 2

Variable	*r*	χ^2	*n*	*df*	*p*	Accuracy	Sensitivity	Specificity
Language group at time 1 (quartile groups for vocabulary and gesture scores at time 1)	0.13	2.36	146	3	.50	86.30%	0.00%	100.00%
Health factors (sex, prematurity, low birth weight, ear infection, visual or hearing impairment, family history, developmental disability)	0.35	17.90	143	6	.006	86.80%	10.50%	98.40%
Demographic factors (maternal education, family income)	0.16	3.62	146	2	.16	86.30%	0.00%	100.00%
Concern at time 1	0.13	2.39	146	1	.12	86.30%	0.00%	100.00%
All variables	0.40	22.87	143	12	.03	86.10%	15.80%	96.80%

Table 4.7:

Standardised canonical discriminant function coefficients of each discriminant function analysis predicting overall and identified concern

Model	Variable	*r* (Overall Concern)	*r* (Identified Concern)
Language group	Productive vocabulary group	1.07	0.66
	Receptive vocabulary group	-0.18	-0.36
	Gesture group	0.08	0.79
Health factors	Sex	0.69	0.67
	Prematurity	-0.25	-0.64
	Low Birth Weight	0.45	0.68
	Ear Infection	0.70	0.60
	Family history	0.22	-0.11
	Visual or hearing impairment	-0.15	0.16
Demographic Factors	Maternal education	-0.48	-0.09
	Household income	1.02	1.02
Concern at time 1	Concern at time 1	1.00	1.00
All variables	Sex	0.36	0.57
	Prematurity	0.04	-0.47
	Low Birth Weight	0.13	0.52
	Ear Infection	0.30	0.45
	Visual or hearing impairment	0.10	0.24
	Family history	-0.01	-0.22
	Maternal education	-0.38	-0.12
	Household income	0.66	0.46
	Concern at time 1	0.08	0.17
	Productive vocabulary group	-0.62	-0.08
	Receptive vocabulary group	0.09	0.10
	Gesture group	0.09	-0.13

Table 4.8:

Results of the discriminant function analysis distinguishing between children whose language did and did not catch up

Variable	*r*	χ^2	*n*	*df*	*p*	Accuracy	Sensitivity	Specificity
All variables	0.57	10.66	35	11	.47	82.90%	54.50%	95.80%

Note: Only 35 children (of the 37 whose language was of concern at some point) had data available for all variables in this analysis

Table 4.9:

Standardised canonical discriminant function coefficients for each variable predicting whether or not children's language caught up after concern was expressed

Model	Variable	*r*
All variables	Sex	-0.53
	Prematurity	0.69
	Low Birth Weight	-0.17
	Ear Infection	-0.15
	Visual or hearing impairment	0.51
	Family history	0.18
	Maternal education	*NA*[1]
	Household income	-0.13
	Concern at time 1	-0.63
	Productive vocabulary group	-0.22
	Receptive vocabulary group	-0.38
	Gesture group	1.03

[1] No children whose language did or did not catch up had low maternal education so it did not contribute to classification accuracy

4.3.6 Exploratory analyses

We ran three exploratory descriptive analyses to investigate why the risk factors were not sensitive enough to correctly classify children whose language was of concern at time 2. Table 4.10 details the number and percentage of children in each Concern group (Overall Concern, No Overall Concern) with each risk factor. We can see from this table that the proportion of children with the risk factor is almost always bigger in the Overall Concern group than the No Overall Concern group. For example, if we consider the family history of language delay or dyslexia risk factor, 20.83% of children in the Overall Concern group have that risk factor, compared to 14.43% in the No Overall Concern group. However, the differences are not big; for most risk factors, a substantial minority of children in the No Overall Concern group also have the risk factor.

Next, we looked at the number and proportion of children with and without overall concern in each of the language quartile groups at time 1. Again, we can see from table 4.11, why language and gesture scores at time 1 do not predict Overall Concern at time 2. Although there are a greater proportion of children in the lowest quartiles who subsequently raise concerns than in the higher quartiles (e.g. 40.82% for 0-25th percentile vs 12.24% in 75-100th percentile for productive vocabulary) the differences are not large or distinct enough to be discriminatory. A substantial minority of children in the higher quartiles go on to develop concerning language, and a substantial minority of children in the lower quartiles do not go on to subsequently raise concerns.

Finally, we created total risk factor scores for each child in the Overall Concern and No Overall Concern groups (see table 4.12 for means and SDs). The total number of risk factors was 10 (being a boy, being premature, having a low birth weight, ear infections family history of language delay or dyslexia, having a developmental disability at time 1, having a visual or hearing impairment at time 1, hearing or communication concerns at time 1, low maternal education and low household income). The mean number of risk factors for children in the Overall Concern group was 1.71 and the mean number of risk factors for children in the No Overall Concern group was 1.04. In addition, the overlap in standard deviation of both groups is quite big, and the range was the same across the two groups (0-5). Therefore, while children in the Overall Concern group experience a slightly larger number of risk factors overall, the differences are not big or distinct enough to be discriminatory.

Table 4.10:

Number and percentage of children with each risk factor split by Overall Concern and No Overall Concern groups

Variable	Have risk factor	Overall concern *n* (%)	No overall concern *n* (%)
Sex (male)	Yes	33(67.35%)	43(44.33%)
	No	16(32.65%)	54(55.67%)
Prematurity	Yes	5(10.20%)	8(8.25%)
	No	44(89.80%)	89(91.75%)
Low birth weight	Yes	5(10.20%)	6(6.19%)
	No	44(89.80%)	91(93.81%)
Ear infection at time 1	Yes	3(6.12%)	0(0.00%)
	No	46(93.88%)	97(100.00%)
Family history	Yes	10(20.83%)	14(14.43%)
	No	38(79.17%)	83(85.57%)
Developmental disability at time 1	Yes	0(0.00%)	0(0.00%)
	No	48(100.00%)	97(100.00%)
Visual impairment at time 1	Yes	1(2.13%)	1(1.03%)
	No	46(97.87%)	96(98.97%)
Concern at time 1	Yes	6(12.24%)	4(4.12%)
	No	43(87.76%)	93(95.88%)
Maternal Education	Yes	3(6.12%)	9(9.28%)
	No	46(93.88%)	88(90.72%)
Family Income	Yes	18(36.73%)	16(16.49%)
	No	31(63.27%)	81(83.51%)

Table 4.11*:*

Distribution of children with and without overall concern across four percentile quartiles for productive vocabulary, receptive vocabulary and gesture scores from 15-18 months

Variable	Concern	0-25th percentile	25th -50th percentile	50th -75th percentile	75th percentile and above
Productive vocabulary	Overall concern	20(40.82%)	17(34.69%)	6(12.24%)	6(12.24%)
	No overall concern	19(19.59%)	13(13.40%)	33(34.02%)	32(32.99%)
Receptive vocabulary	Overall concern	14(28.57%)	13(26.53%)	12(24.49%)	10(20.41%)
	No overall concern	19(19.59%)	19(19.59%)	20(20.62%)	39(40.21%)
Gesture scores	Overall concern	15(30.61%)	13(26.53%)	8(16.33%)	13(26.53%)
	No overall concern	15(15.46%)	23(23.71%)	35(36.08%)	24(24.74%)

Table 4.12*:*

Mean and standard deviation of the number of risk factors for children with and without overall concern. Minimum and maximum number of risk factors experienced by any one child in each group is also listed.

	M	*SD*	Minimum	Maximum
Overall concern	1.71	1.14	0	5
No overall concern	1.04	1.09	0	5

4.4 Discussion

The goal of the present study was to investigate if we could use a combination of risk factors, earlier vocabulary and gesture scores, as well as early parental concern for language development to predict later concern for language development in children who took part in the UK-CDI Project.

We first investigated if there was a difference in the demographics of families who did and did not respond to our questionnaire. We found that parents from families with higher income and higher maternal education were more likely to respond to the follow-up questionnaire than parents from families with lower income and lower maternal education. This is a common problem when collecting data from lower SES families (Reilly et al., 2010). However, when we consider that DLD affects 7.58% of all children (Norbury et al., 2016), and that 13.70% of our sample had an identified disability, we believe we have still collected data from a representative proportion of the population, at least when it comes to language ability, if not socio-economic status. It is however, important that low SES families are represented in this area of research. Future research should therefore make an increased effort to contact families represented in lower SES brackets. We may have been more successful at encouraging families to participate if we had personally contacted them, either via email or phone.

We then examined the strength of parental concern over time. Contrary to our predictions, overall concern at time 2 was not predicted by parental concern at time 1. This is due to there being so few parents reporting concern at time 1; only 6 parents of children who expressed concern at time 2 also expressed concern at time 1 (see table 4.10). Therefore, we are hesitant to conclude that, where more parents express early concern for their children's language, there would be no relationship between parental concern at time 1 and Overall Concern at time 2.

Following this, we used discriminant function analyses to examine if different combinations of health and demographic risk factors as well as earlier vocabulary and gesture scores could discriminate children for whom there was, and was not, concern for their language development. We first examined the role of earlier language and gesture scores in predicting later concern for language impairment. These variables did not successfully discriminate between children with and without concern for their language development due to poor sensitivity; they failed to correctly classify the children in the Overall Concern group. When we examine the number of children in each vocabulary and gesture quartile (table 4.11), we see that children with and without overall concern were distributed across all four quartiles with very little clustering at each end for each group. To be expected, children for whom there was overall concern were less likely to be represented in the top two quartiles for productive vocabulary. However, they were distributed across the bottom two quartiles (0-25th and 25th-50th) and, while their productive vocabulary scores were below average (below the 50th percentile), they were not very low. In addition, 19.59% and 13.40% of children for whom no one expressed concern for their language development at time 2 were also in the bottom 0-25th and 25th-50th respectively at time 1. This means that early language and gesture ability did not reliably distinguish between children whose language was, and was not, of later concern.

One reason that could explain why early vocabulary and gesture scores, recorded at 15-18 months, do not predict concern for language development at 4-6 years, is a disconnect between the skills that children are required to master early on to acquire language (i.e. to learn to use and interpret gestures and words) and the skills that characterise later acquisition (syntax and pragmatics). Although some have argued that vocabulary and syntax development are strongly correlated (Bates & Goodman, 1999), others have argued strongly that syntactic and vocabulary acquisition are governed by different learning mechanisms (innate syntactic ability; e.g. Valian, 2014). Language impairments in older children tend to be characterised by greater difficulties in syntax or pragmatics (Bishop, 2014; Norbury & Bishop, 2002) than in vocabulary (Bishop & Snowling, 2004; Leonard, 2014). Therefore, one possible reason for these results is that different acquisition mechanisms underlie vocabulary and syntax acquisition, and, thus, that delays in vocabulary and syntax acquisition stem from different causes (see van der Lely, Rosen, & McClelland, 1998, for a theory of specific language impairment based on this premise).

Alternatively, it is possible that the change in children's environments between 15-18 months and 4-6 years is having an effect. Previous research has shown that once children start attending playgroups and nurseries (typically at about 2-3 years of age), this has a substantial impact on their cognitive development (Sylva, Stein, Leach, Barnes, & Malmberg, 2011; Turner, 1974). It might be that the environmental influences that affect language development early in life are now no longer influencing later language, and vice versa. For example, children who hear little language in the home early in life, and thus develop slowly at first, might start to attend a nursery that promotes language development, and thus start to thrive. In other words, changes in children's environment over the preschool years may break the link between early and late language development.

We then examined if a number of health or demographic risk factors could predict later concern for language impairment. As with early language skills above, these risk factors did not successfully identify children for whom there was concern for their language development due to a poor sensitivity. Although the Overall Concern group experienced, on average, a greater number of risk factors (see table 4.12: 1.71 vs. 1.04), there was lot of overlap (wide and overlapping standard deviations and ranges) in the number of risk factors in each group. This means that no combination of health or demographic risk factors was experienced almost exclusively by children with concern for their language development and therefore no factors were discriminant enough to distinguish between these two groups of children. Furthermore, when we consider that there were a maximum of 10 risk factors and the most risk factors any one child experienced was 5, we can see that neither children with nor children without overall concern for their language development were exposed to a very high number of these risk factors. This result is consistent with previous research, which has shown that health and demographic risk factors are better at predicting individual differences than they are at predicting language impairment or concern for language development (Harrison & McLeod, 2010; Reilly et al., 2010). For example, Harrison and McLeod (2010) found health and demographic risk factors did not predict parental concern for vocabulary development or use of speech-language pathology services, with a combination of these factors yielding poor levels of sensitivity.

Following this, we examined if early parental concern for language development predicted later overall concern for language development. Again, this did not improve our ability to discriminate between children with and without concern due to poor sensitivity. Previous research has shown that parental concern for language delay can benefit clinical detection of language impairment (Glascoe, 1991; Glascoe & Dworkin, 1995). However, this research has been typically run with older children. The findings here suggest that very early concern for language development may not be as beneficial for predicting later problems. Again, it is possible that the factors driving parental concern at time 1 and time 2 are not the same. When children are 15-18 months old, delay in productive vocabulary development is responsible for parental concern. However, by the time children are 4-6 years old, vocabulary is typically no longer a cause for concern. As mentioned above children experiencing language impairment at 4-6 years typically present difficulties associated with syntax acquisition (Bishop & Snowling, 2004; Leonard, 1998; Leonard, 2014). Therefore, parental concern at time 1 and time 2 may stem from different sources.

Similarly, with all of the factors combined, while we were successfully able to distinguish between children in terms of accuracy and specificity, the sensitivity of each of these models was low. This means that these models all failed to identify those children for whom there is concern for their language development. This is particularly surprising when we consider that this full model includes all risk factors, earlier vocabulary and gesture scores and early concern for language development. However, the finding is in line with previous research. Reilley et al (2010) found that while risk factors were successful in predicting continuous language scores at 4 years (i.e. individual differences), they were unable to correctly classify children with specific language impairment (now DLD). Looking at our exploratory analyses, we can begin to understand these findings; there were no risk factors that consistently, and distinctly, impacted children for whom there was later concern.

One possible explanation for why risk factors, early vocabulary and gesture scores, and early concern for language development do not predict later concern for language development is that our measure of later concern was not reliable. It was possible that the inclusion of parental concern for language development and identification by the Healthy Child Programme's 2 Year Review added too much noise. However, we controlled for this possibility by repeating the analyses using identified disability. Still, these variables did not successfully predict identified disability and thus, it was not the reliability of our concern measure driving these findings.

Finally, we investigated if catch-up in language delay is associated with exposure to fewer risk factors and better scores on earlier vocabulary and gestures. In line with the results of the previous models, this model had poor sensitivity and did not correctly classify children whose language did not catch up - children whose language development was of the greatest concern. Again, we find that risk factors as well as earlier vocabulary and gesture scores are not exclusively impacting children with the poorest language development and are therefore, not predictive of persisting language impairment.

Throughout all of the analyses, there was a failure to identify children with overall concern for their language development, with identified disability and whose language did not catch up due to a consistent lack of sensitivity. This is in line with previous research; it is very difficult to classify children whose language is of most concern over time (Reilly et al., 2010). We know that a substantial proportion of children who present with early delays in language acquisition continue to perform poorly on measures of language and literacy skills later in childhood (Rescorla, 2002, 2009). However, these measures are not accurate enough to classify children later in life. Therefore, more research is required to identify predictors of language delay over time.

4.5 Conclusion

The present findings shed light on the role of health and demographic risk factors in predicting later language outcomes. Risk factors are currently recommended as a starting point when monitoring language delay in young children because the incidence of persisting language impairments is greater in children from families exposed to greater risks (Harrison & McLeod, 2010). However, since we have found that these risk factors, alone and combined, do not allow us to accurately predict concern for children's language development over time we would not recommend that they be used when screening children at risk for language impairment on the basis of a delay in productive vocabulary acquisition.

5 Chapter 5: Discussion

There were three main aims of this thesis. The first aim was to identify predictors of individual differences in vocabulary acquisition. The second aim was to establish if the same predictors can be used to identify children who were experiencing a delay in productive vocabulary development. The third aim was to investigate if combinations of early risk factors and language skills can identify children for whom there is concern about their language development between 4-6 years. Chapter 2 identified factors implicated in predicting individual differences in productive vocabulary development at 24 months, and in distinguishing between typically developing children and children who were slow to talk at 24 months. However, nearly all of the children studied in chapter 2 had language development within the typical range; few were so delayed as to be classified as late talkers. Thus, in chapter 3 we collected an additional sample of late talking children. We examined predictors of individual differences in vocabulary development at 25 months in this combined sample of typically developing and late talking children, and the ability of these predictors to discriminate between typically developing and late talking children. Finally, chapter 4 extended the investigation to the role of early language skills in later language impairments, by examining the relationships early language skills, early concern for language development, risk factors and later concern for language development.

This discussion chapter will first summarise the findings from the three empirical chapters. Following this, it will provide an overview of the results for each predictor variable, in the light of the findings from all chapters and the previous literature. Next, the chapter will present a summary of the discriminant function analyses throughout the thesis and their implications for how we use combinations of predictor variables to discriminate between different groups of children in terms of their language ability. Finally, directions for future research will be presented.

5.1 Chapter 2: Individual differences in productive vocabulary: Identifying children who are slow to talk.

In this chapter, we investigated factors that contributed to individual differences in vocabulary development and established if these factors, individually and combined, could be used to distinguish between typically developing children who were faster, and slower, to talk at 24 months. The aim of this was to investigate potential factors that could then be used to identify late talking children who are traditionally identified around 2 years old. We found that conversational turn counts, earlier receptive vocabulary, mean length of utterances (MLU) non-word repetition (NWR) and speed of linguistic processing, all predicted unique variance in productive vocabulary at 24 months after controlling for earlier productive vocabulary and child sex. We controlled for these factors for two reason. First, they have been shown to be robust predictors of individual differences in later productive vocabulary and are relatively easy to assess; if these factors can explain the variance in productive vocabulary alone, there would be no need to run additional assessments. Second, we were interested in establishing what unique variance the additional factors explained in order to have a better understanding of their unique role in language development, and in turn, their usefulness in identifying late talking children.

When we examined the discriminatory ability of each of these factors individually, we found that they were successful in distinguishing between the children who were, and were not, slower to learn to talk. In addition, when combined, all of these factors were successful in discriminating between the two groups of children. When using only the non-experimental factors or factors collected from earlier data points, they could still successfully discriminate between these two groups of children. However, when using only earlier non-experimental measures, the factors were less successful in identifying children who were slow to talk. Therefore, when using earlier measures to predict group membership at 24 months, the inclusion of experimental measures (speed of processing at 19 months) improved our ability to discriminate between typically developing children and children who were slow to talk.

5.2 Chapter 3: Individual differences in vocabulary acquisition in late talking and typically developing children.

The research outlined in the second empirical chapter replicated the above research with an extended sample of late talking and typically developing children. Chapter 3 identified whether the same factors which were successful in predicting individual differences in a sample of typically developing children, were also predictive of differences between typically developing and late talking children. Children's earlier receptive vocabulary, the mean length of their three longest utterances (M3L), earlier gesture scores, and NWR all predicted unique variance in productive vocabulary scores at 25 months, after controlling for sex and earlier productive vocabulary scores. Next we assessed the ability of these variables to distinguish between late talking and typically developing children. Late talkers were classified using two commonly applied criteria. First, children who were not producing 50 words were classified as late talkers (50 word cut off). Second, children who were not combining any words together were classified as late talkers (failure to combine cut off). Using the 50 word cut off, earlier productive and receptive vocabulary, concurrent M3L and NWR could all successfully distinguish between late talking and typically developing children. However, when all variables were combined, only NWR made a unique contribution to group membership. The earlier non-experimental measures alone (sex, earlier productive vocabulary, earlier receptive vocabulary and earlier gestures) did not successfully discriminate between late talking and typically developing children, with a failure to identify any late talking children. Using the failure to combine cut off, earlier gestures and concurrent NWR successfully distinguished between late taking and typically developing children. Again, with all variables combined, NWR contributed the most to successfully distinguishing between these two groups of children. However, the NWR task presented a limitation due to sample size. Most late talking children did not complete the NWR task and therefore, the resulting sample size was substantially reduced. Using only earlier non-experimental measures, again, this combination of variables did not successfully distinguish between typically developing and late talking children. However, the sensitivity of this combination of variables was better than that using the 50 word cut off.

In summary, the results from chapter 3 demonstrate that early measures of language development can be used to predict individual differences in vocabulary development at 25 months. In addition, these measures can be used individually to distinguish between late talking and typically developing children. However, when combined, the measures lose their ability to discriminate between these two groups due to a lack of sensitivity.

5.3 Chapter 4: Follow-up study from the UK-CDI Project: Can risk factors, vocabulary skills and gesture scores predict later concern for language development?

This chapter presented a study following up children who had taken part in the UK-CDI Project at 15-18 months of age, and were now aged between 4-6 years. The purpose of this was to identify if any concern about their language had ever been expressed, and to examine if we could predict this concern from factors recorded when children were 15-18 months old. The aim of this study was to identify if different combinations of language skills, risk factors and early concern for language development could reliably improve the accuracy of predicting later concern for language development.

Whether we examined factors in individual groups or combined together, none of our factors - language skills, health risk factors, demographic risk factors and early concern for language development - successfully discriminated between children for whom there was and was not concern for their language development. However, there was a possibility that our measure of concern was too inclusive, so we repeated these analyses with a stricter criteria – identifying only children who had an identified disability that could impact their language development over time (developmental disability, diagnosis of a language disorder, and/or having a hearing or visual impairment). Again, the different combination of factors, individually and combined, did not successfully identify children.

We then examined whether or not language skills from 15-18 months, health and demographic risk factors, and early concern for language development could distinguish between children whose language did or did not catch up after concern for their language development had been identified. Again, we found that these variables did not successfully identify children whose language did not catch up.

In summary, the results from chapter 4 suggest that none of our early predictor variables allowed us to identify the children of greatest concern: children with concern for their language development, children with an identified disability and children whose language never caught up. We concluded that this is most likely due two possible, not mutually exclusive, reasons. First, risk factors are not discriminant enough to distinguish between groups of children; there is too much overlap between children with and without concern for their language development. Second, there may be a disconnect between the development of early language skills (vocabulary and gesture acquisition) and later language development (syntax and pragmatics acquisition).

5.4 Overview of each predictor variable: Interpretations of their contributions to the results

Given the findings summarised above, what implications do these results have, overall, for our understanding of the causes of individual differences, and of language difficulties, in early childhood? In the following section, we discuss each predictor variable in turn, collating the evidence across chapters, and discuss its contribution to language development.

5.4.1 Child sex

Child sex predicted individual differences in productive vocabulary at 24 and 25 months in chapters 2 and 3. This is in line with previous research that robustly demonstrates that girls tend to more advanced in their language development than boys (Fenson et al., 1994; Horwitz et al., 2003).

However, throughout all three empirical chapters, child sex contributed very little to the identification of children whose language development was slow. In chapter 2, child sex did not contribute to our ability to discriminate between children who were and were not slow to talk in our typically developing sample. In chapter 3, again, child sex did not contribute to discriminating between late talking and typically developing children using concurrent and earlier measures; the overall sensitivity of these models was poor. Similarly, in chapter 4, while child sex contributed a substantial amount to the classification accuracy of identifying children for whom there was concern, the overall sensitivity of these models was poor and thus, child sex did not contribute to the overall success in identifying children whose language was of concern. Finally, in models using a combination of variables to discriminate between children whose language did and did not catch up, child sex did not improve this model's ability to identify children whose language did not catch up.

We can see from these studies that child sex is a strong predictor of individual differences in vocabulary and we know from previous research that is in implicated in language impairment (Reilly et al., 2010; Rescorla, 2011). However, in the research presented here, sex was not a reliable factor for distinguishing between children who do and do not have typically developing language and fails to identify the children whose language is of most concern. The reason we find individual differences when using sex as a predictor variable is because there are slightly more girls at the top of the scale and slightly more boy at the bottom. However, given that only a small proportion of the population have language delays (Norbury et al., 2016; Tomblin, Records, Buckwalter, Zhang, Smith & O'Brien,1997), most boys will not experience a delay in language acquisition. Therefore, sex is a very unreliable predictor of language delay.

What is most surprising is that, in all three chapters, sex did not add much discriminatory power when combined with other measures. Although our results show that sex is not a good discriminator on its own, we might expect it to add some accuracy when combined to other measures. However, sex was such an unreliable discriminator that it did not actually improve classification accuracy. Thus, although some advocate the use of sex as a risk factor, our studies suggest that there is no reason for its inclusion because it adds no additional explanatory power.

5.4.2 Family history of speech or language impairments

In both chapter 2 and chapter 3, family history did not correlate with vocabulary and thus its discriminatory ability was not assessed in detail. One exploratory analysis, outlined in chapter 2, included family history along with five other measures (child sex, earlier productive vocabulary, earlier receptive vocabulary, MLU and NWR) but it did not increase the model's accuracy. In addition, when we examine the descriptive statistics of the children in chapter 3, we see that rates of family history of speech or language impairment was not greater for children in the low language sample.

In chapter 4, family history of speech or language impairment was used, along with other demographic factors, to predict later concern for language development between 4-6 years. Here, family history did not contribute to classifying children with later concern for their language impairment or children with an identified disability that could impact their language development. When we examine the exploratory descriptive analyses, it can be seen that family history of speech or language impairments was a poor classifier because, although the proportion of children with family history was higher for children with concern for their language development, there was still too much overlap between the two groups of children.

The finding that, throughout these chapters, family history of speech or language impairments did not relate to vocabulary or predict later concern for language development is surprising and contradicts previous research (Bishop et al., 2012; Reilly et al., 2007). The finding that family history does not correlate with vocabulary is likely due to the small number of children with a family history of speech of language impairments in our samples. However, the finding that family history of speech or language impairment does not predict later concern for language development is harder to explain. It is possible that although family history is implicated in language development, it is, like child sex, a poor discriminator. The exploratory descriptive analyses in chapter 4 show that while there may slightly be more children with a family history at the bottom end of the distribution, and thus have concern for their language development, most children with and without a family history have no concern for their language development. In summary, we know, from the literature, that family history of speech or language impairments does have an impact on language development and impairment (Bishop et al., 2012; Reilly et al., 2007). However, its contribution to language impairment was not strong enough for it to be good at discriminating between children with and without early vocabulary delay and children who do and do not have concern for their language impairment.

5.4.3 Early parental concern for language impairment

In chapter 4, we examined if early parental concern for language development, early language and gesture skills, as well as health and demographic risk factors, could distinguish between children with and without concern for language development, with and without an identified disability, and whose language did and did not catch up. Alone, early parental concern for language impairment had no effect on the likelihood of later concern. In addition, it did not correctly classify any children of concern or with an identified disability. When combined with early language and gesture skills, as well as health and demographic risk factors, early parental concern for language impairment contributed very little to classification accuracy. Similarly, when distinguishing between children whose language did and did not catch up, early concern, again, did not contribute to classification accuracy.

Overall, early parental concern for language development was not sensitive enough to improve predictions about later concern for language development. While there is research showing the benefit of parental concern for identifying children at risk for developing language impairment, this has typically been done with older children (Glascoe, Altemeier, & MacLean, 1989; Glascoe & Dworkin, 1995). In our study, rate of parental concern at 15-18 months was low for both children with and without concern for their language development, and thus, it was not a useful discriminator this early in childhood.

Alternatively, it may be that the factors driving concern for language development between 15-18 months are not the same as those at 4-6 years. For children with an early delay in language development, vocabulary acquisition is often the source of parental concern for language development (Rescorla, 2011). However, for most children with persisting language impairments, vocabulary will not be the source of parental concern; children with persisting language impairments usually present difficulties with syntax (Bishop & Snowling, 2004; Leonard, 1998) or pragmatics (Bishop, 2014; Norbury & Bishop, 2002). As mentioned in chapter 4, later concern for language development may be driven by children's syntax or pragmatic skills. Bates and Goodman (1997) argue that language and syntax are strongly linked, and the best predictor of syntax acquisition is a child's own vocabulary skills. They suggest that there is a bidirectional relationship between vocabulary and grammar, once children can produce a certain number of words, and that this does not dissociate throughout development. The results from chapter 2 and 3 support this view, showing that vocabulary and grammar are related up until children are, at least, 30 months old. However, the results from chapter 4 suggest that sometime after 30 months, this relationship begins to dissociate. For example, Locke (1997) proposed that there are specialised phases of language development that are time-locked; subsequent phases cannot be activated if the previous phase is delayed. Following this theory, syntax will not develop if the lexicon size is insufficient. Locke (1997) argued that each phase is mediated by distinct neural mechanisms. Therefore, vocabulary size determines the activation of grammar but not the success of grammar. It is, thus, possible that once children have a certain understanding of the rules of syntax for their language, vocabulary is no longer important for the success of syntax development. If this were the case, it would explain why early vocabulary skills do not predict later concern for language development, where concern is a result of a delay in syntax development.

5.4.4 Distal risk factors: Prematurity, low birth weight, ear infection, hearing or visual impairment, developmental disability, maternal education, family income

In chapter 4, we examined the contribution of health and demographic risk factors in distinguishing between children with and without concern for language development, with and without an identified disability, and whose language did and did not catch up.

We consider these factors to be distal because their effect on language is indirect. Previous research has shown that these factors do impact language development over time (Harrison & McLeod, 2010; Reilly et al, 2010), however, they have to be having an effect via proximal factors. For example, Jansson-Verkasalo et al (2004) found that the prevalence of birth complications is higher in children who are slow to talk. Since birth complications cannot be directly influencing language acquisition, they must impact some cognitive mechanisms responsible for language development and in turn impact language acquisition.

Alone, health risk factors (sex, prematurity, low birth weight, ear infection, hearing or visual impairment, family history of speech or language impairment and developmental disability) failed to identify children with concern and children with an identified disability due to poor sensitivity. While the presence of an ear infection contributed the most to the classification accuracy of these variables, overall the sensitivity of the model was still poor. Demographic risk factors (family income and maternal education) also failed to identify children with concern and children with an identified disability due to poor sensitivity. When these factors were all combined, along with earlier language and gesture skills, and earlier concern for language impairment, all factors still failed to distinguish between these children, again, due to a lack of sensitivity.

As with sex and family history of speech and language impairment above, we know from previous research that these factors have an impact on language development and are implicated in language impairment (Harrison & McLeod, 2010; Reilly et al., 2010). Like child sex and family history of speech or language impairment, these factors were not strong discriminators. The exploratory descriptive statistics in chapter 4 show that risk factors tend to impact children with concern for their language development slightly more than those without concern. However, this difference is marginal and still, the majority of children with and without concern for their language development experience no risk factors. It is, however, surprising that when combined, these factors did not improve classification accuracy. Like child sex, they were such poor discriminators that even additively, their accuracy did not improve. Thus, our studies show that such risk factors cannot be used to predict later concern for language impairment due to poor sensitivity.

5.4.5 Environmental measures: Adult word count and conversational turn count measures using LENA

In chapter 2, we found that adult word count did not correlate with productive vocabulary at 24 months. While conversational turn counts did correlate, it did not predict individual differences in productive vocabulary after controlling for sex and earlier productive vocabulary. In chapter 3, neither of these measures correlated with productive vocabulary at 25 months and thus, their discriminatory ability was not assessed. The finding that adult word count does not correlate with vocabulary development is inconsistent with previous research. There is a wealth of research demonstrating the role of input on vocabulary development (Hurtado, Marchman, & Fernald, 2008). However, adult word count, as captured by the LENA (LENA Research Foundation, 2014), includes all adult vocalisations within the child's environment and does not distinguish between child-directed speech and overheard speech. Recent research has shown that it is the quality of input, rather than the quantity of input, that influences children's vocabulary development (Cartmill et al., 2013; Jones & Rowland, 2017; Rowe, 2012). Thus, it is likely that this measure of adult word count captured too much overheard speech that was not impacting children's vocabulary development. A measure of adult word count that includes only child-directed speech could provide a better assessment of its unique role in vocabulary development and language delay.

Conversational turn count is a measure of the amount of dialogue between children and adults. The number of turns taken is determined both by the adult speech and how much language the child can produce. Children who know more words will take more turns. Therefore, it is crucial to control for children's earlier productive vocabulary if we want to examine whether or not turn-taking has an effect on language development over time. While conversational turn counts did correlate with productive vocabulary at 24 months in chapter 2, it failed to predict unique variance in productive vocabulary over and above earlier productive vocabulary and child sex. This finding is inconsistent with previous research. Zimmerman et al. (2009) found conversational turns predicted later performance on a preschool language assessment even after controlling for earlier performance on the same assessment, suggesting that conversational turns do have a unique contribution to language development. However, there is still very little research on conversational turns on language development, and most of it still does not control for previous vocabulary (Gilkerson et al., 2017; Romeo et al., 2018). Therefore, we are still unsure if conversational turn count is an important predictor of language acquisition and what role they may play in language impairment.

5.4.6 Earlier productive and receptive vocabulary

The results from chapters 2 and 3 show us that receptive vocabulary plays a unique role in productive vocabulary development; the words that children understand have an impact on the words they learn over and above the words they already say. There is a wealth of research showing that there is individual variation in the words that children know, but our results suggest that there is also individual variation in the gap between the number of words children understand and the number of words they say; some children understand a lot of words and hardly say any of them, while other children understand a lot of words and say almost all of them. Our findings suggest that there is a unique role played by receptive vocabulary, over and above productive vocabulary; children who understand a lot of words- even if they are not saying them- will learn more subsequent words compared to children who understand only a few words- even if they are saying all of those words.

Throughout chapter 2 and 3, both earlier productive and receptive vocabulary were two of the best discriminators; successfully distinguishing between children who were and were not slow to talk, and between late talking and typically developing children using the 50 word cut off. However, these measures were not strong enough to discriminate between children with and without concern for their language development at 4-6 years. While earlier productive and receptive vocabulary successfully discriminated between children who were and were not slow to talk up to 30 months, this relationship breaks down sometime between 30 months and 4 years. Thus, earlier language scores are not sensitive enough to classify children at 4-6 years.

At first sight these results seem non-intuitive; why would there be a disconnect between language acquisition in the very early years (up to 2;6) and acquisition later in childhood (between 4 and 6 years). However, these results fit with much of the literature, which also finds that it is extremely difficult to classify children who have concern for their language development between 4-6 years using language measures taken between 15-30 months (Bishop et al., 2012; Reilly et al., 2010).

There are a number of possible reasons to explain why early vocabulary scores are poor at identifying children who have concern for their language development later in childhood. First, it is possible that our measure of later concern for language development was not reliable enough to capture true language impairment. However, to control for this, we also looked at identified disability using three criteria known to impact language impairment: developmental disability, diagnosis of a language disorder, and/or having a hearing or visual impairment. Still, early vocabulary scores were unable to distinguish between children with and without an identified disability. Therefore, it was not the reliability of our concern measure leading to these results; parents were not reporting concern for their children’s language development when there was nothing wrong with their language.

Second, it is possible that different factors affect language development in the early and the later preschool years. Over these years, other factors come into play; children’s environments change a lot as they start to go to nursery and school. For example, some research shows that nursery (Sylva, Stein, Leach, Barnes, & Malmberg, 2011) and playgroup attendance (Turner, 1974) both influence cognitive and language development during preschool years. For example, Sylva et al (2011) found that the quality and quantity of non-maternal group care was associated with higher cognitive scores on the Bayley Scales of Infant Development at 18 months. All of this variation will impact children’s language development and thus, it is very hard to classify these children over time. Our results are in line with this research; we have good classification accuracy using these measures up until 30 months, but at 4-6 years, they are no longer reliable.

Third, and most interesting, is the likelihood that the factors implicated in concern at 4-6 years are no longer strongly related to vocabulary development.

Children with persisting language impairment, into later childhood, present difficulties with syntax (Bishop & Snowling, 2004; Leonard, 2014) or pragmatics (Bishop, 2014; Norbury & Bishop, 2002).. As discussed above, Bates and Goodman (1997) propose that vocabulary and syntax are strongly related in early language development and that these two factors remain closely associated throughout life. However, Locke (1997) proposes that the development of the lexicon and syntax emerge from separate linguistic mechanisms that activate in sequence; when the lexicon is large enough, syntax development is activated, but the success of syntax development is not necessarily dependent on any further development of the lexicon. Following Bates and Goodman's (1997) theory, we would expect that there would be a relationship between vocabulary and later concern for language development even if concern for language development is driven by delays in syntax development. However, following Locke's (1997) theory, if vocabulary and syntax stem from separate linguistic mechanisms, the relationship between vocabulary and syntax would not be strong enough to predict later concern for language development. Examining both theories, we might be able to explain why vocabulary and syntax are strongly related in chapters 2 and 3 but not in chapter 4. It is highly likely the vocabulary and syntax are linked in early language development. However, it is possible that in later language development these two factors dissociate. If this is true, the reason we find such poor predictive classification accuracy later in childhood is because these two factors are now stemming from different mechanisms. Therefore, we would not expect to find any relationship between early vocabulary development and later concern for language development on the basis of delays in syntax development.

It is also possible, that syntax and vocabulary development separate earlier in childhood than we would expect. Our findings in chapter 3 support this idea. When late talking children were classified on the basis of word combination, a measure of syntax, productive and receptive vocabulary were not sensitive enough to identify late talking children. It is possible that there is a difference between children who are slow to produce words and children who are slow to use syntax, and it is those syntax problems that would be identified as a matter of concern at 4-6 years.

5.4.7 Mean length of utterance

Mean length of utterance (MLU) and the mean length of a child's three longest utterances (M3L) are measures of productive syntactic ability. In chapter 2, MLU at 24 months was successful in predicting unique variance in productive vocabulary development at 24 months and in discriminating between children who were and were not slow to talk. Similarly, in chapter 3, M3L at 25 months predicted unique variance in productive vocabulary at 25 months, remained significant in the full regression model with five other measures, and successfully discriminated between late talking and typically developing children. Previous research has shown a strong relationship between vocabulary and syntax acquisition; as the lexicon grows, it provides the foundations for syntax to develop (Bates & Goodman, 1999; Marchman & Bates, 1994).

Other research however, proposes that, vocabulary and grammar emerge in parallel and as children's understanding of grammar grows, their vocabulary benefits from this and in turn, grows further (Bates & Goodman, 1997; Dixon & Marchman, 2007) . Dionne, Dale, Boivin and Plomin (2003) found that the syntax patterns learned from the lexicon aid further vocabulary acquisition. It is therefore possible, that at around 2 years, knowledge of syntax is crucial for vocabulary development. Many children experience an explosion in vocabulary between 18-24 months (Goldfield & Reznick, 1990). At the same age, children begin to acquire syntax (Bates et al., 1994). It is possible that syntax, at least in part, contributes to this explosion of vocabulary and thus, if children are experiencing a delay in syntax acquisition, it may, for a time, present as a delay in vocabulary acquisition. The findings in chapters 2 and 3 support this idea; in chapter 2 and 3, we found that syntax predicted variance in vocabulary at both 24 and 25 months suggesting that, at this age, syntax and vocabulary development are tightly related. In addition, in chapter 3, syntax classified both typically developing and late talking children using the 50 word cut off. However, the finding that vocabulary measures did not correctly classify late talkers when using a syntax related cut off (failure to combine) suggest that, at least at this age, it is syntax impacting vocabulary development.

Similarly, in chapter 4, our early measures of vocabulary- again from 15-18 months- were not successful in predicting later concern for vocabulary development. When we consider that children who have persisting language impairment, often present with impairments related to syntax, these findings begin to make sense. While we know that there is a relationship between vocabulary and syntax early on, it is possible, as mentioned above, that these factors diverge as children are getting older. Therefore, by the time children are 4-6 years old, the complexity of their grammar is no longer a reflection of the number of words they know but of something other than vocabulary. If there is a disconnect between vocabulary and syntax, it is possible that an early measure of syntax would be a better indicator of later language impairment. While syntax cannot be measured as early as 15-18 months, a measure from 24-30 months might improve our ability to distinguish between children with and without concern for their language development.

5.4.8 Early gesture use

Early gesture skills were successful at discriminating between typically developing children and late talkers using the failure to combine cut off (chapter 3). However, early gesture skills were not successful at predicting individual differences in typically developing children after controlling for child sex and earlier productive vocabulary (chapter 2) or at discriminating between children with and without later concern for their language development (chapter 4). Of particular interest is the finding that when classifying late talkers using the failure to combine classification, earlier gesture use was one of the most successful in discriminating between typically developing and late talking children. Previous research shows that gestures impact vocabulary development (Hsu & Iyer, 2016), yet in chapter 2, gestures did not predict unique variance in productive vocabulary development. It is possible that early gestures play a role in very early vocabulary development and when vocabulary is larger, the relationship between gesture and vocabulary weakens. This would explain why we find an effect of gestures with a sample that includes late talking children and not with a sample of typically developing children.

It is also possible that we have a floor effect of productive and receptive vocabulary between 15-18 months within the sample of late talking children and therefore, the best measure of variation was their gesture production. Gesture use is a measure of very early communicative abilities (Colonnesi, Stams, Koster, & Noom, 2010; Iverson & Goldin-Meadow, 2005) and thus, when vocabulary scores are low, gestures may become the best indicator for communicative ability. The results in chapter 3 support this idea; when we discriminate children on the basis of their very early syntax skills (combining words), early gesture skills, and not productive and receptive vocabulary, could successfully discriminate between late talking and typically developing children.

5.4.9 Speed of linguistic processing

Speed of processing at 19 months predicted unique variance in productive vocabulary in chapter 2. When discriminating between typically developing children and children who were slow to talk, speed of processing was less successful at identifying typically developing children using the 25th percentile cut off. However, when using a combination of variables recorded before 24 and 30 months, speed of processing improved the classification accuracy. Interestingly, in chapter 3, speed of processing at 25 months did not correlate with productive vocabulary at 25 months. However, this is in line with recent research (Peter et al., 2019) where speed of processing at 25 months did not correlate with productive vocabulary between 18-36 months. Speed of processing on adjective-noun trials did, however, correlate with productive vocabulary at 25 months. This is most likely because of an increase in variance in children's reaction time due to the inclusion of adjectives. Adjectives are acquired later than nouns (Mintz & Gleitman, 2002) and thus, may not have been as high frequency for all children. However, speed of processing on adjective-noun trials did not explain unique variance in productive vocabulary at 25 months, suggesting that the role of speed of processing in concurrent vocabulary development is mediated by both earlier productive vocabulary and child sex.

If speed of processing is a measure of vocabulary size, then we would not expect it to predict later concern for language development since the results in chapter 4 demonstrated no relationship between early vocabulary and later concern for language development. However, previous research has shown a relationship between speed of linguistic processing and syntax acquisition (Fernald, Perfors, & Marchman, 2006; Peter et al., 2019). Peter et al. (2019) found speed of processing at 19 months predicted growth in syntax from 19-30 months, even after controlling for productive vocabulary at 19 months. When we consider that children with later language impairments often have difficulties with syntax (Bishop & Snowling, 2004), it is therefore, possible that speed of processing would be a good discriminator for children with and without concern for their language development later in childhood on the basis of a delay in syntax development. For this reason, in chapter 3, we examined the role of speed of processing on adjective-noun trials in discriminating between late talking and typically developing children (late talking as defined by the failure to combine classification). We found the sensitivity and specificity of this model did not reach significance. However, this is most likely due to there being so few late talkers with speed of processing scores on adjective-noun trials. Further research on late talking as determined by a delay in syntax acquisition could help disentangle to role of speed of processing in early language delay.

5.4.10 Non-word repetition

Throughout chapter 2 and 3, NWR was a robust predictor of individual differences in productive vocabulary development and was successful in discriminating between typically developing children and children who were slow to talk (chapter 2), and typically developing and late talking children (chapter 3). Jones, Gobet and Pine (2007) propose that NWR is a proxy measure of vocabulary size; they argue that performance on NWR tasks increases with age simply because of an increase in the vocabulary size. However, our finding that NWR explained variance in later productive vocabulary scores after controlling for earlier vocabulary suggests that it is not just a measure of vocabulary, but that it plays a unique role in vocabulary development.

NWR is a clinical marker for language impairment (Botting & Conti-Ramsden, 2001; Conti-Ramsden, Botting, & Faragher, 2001). It is widely accepted that children with language impairment struggle with syntax development (Bishop & Snowling, 2004; Leonard, 1998). Phonological working memory plays a role in syntax acquisition, in that children who perform better on measures of phonological working memory, produce longer, syntactically richer sentences (Adams & Gathercole, 2000; Blake, Austin, Cannon, Lisus, & Vaughan, 1991). Therefore, we would expect NWR performance to predict later concern for language impairment when concern for language come from children’s syntax development. However, it is impractical to run NWR tasks with children younger than two. In both chapter 2 and chapter 3, most of the children who did not complete the NWR task were those who had the lowest productive vocabulary scores. Therefore, although NWR performance might be a robust predictor in principle, in practice it is difficult to get NWR scores from the children whose language development is of the most concern.

In summary, earlier and concurrent measures of language proficiency (productive and receptive vocabulary, gestures and MLU or M3L), and NWR are the most promising predictors of vocabulary at 2 years and can successfully discriminate between typically developing children and late talking children. However, when predicting later concern for language development, no individual group of factors, or overall combination of factors could, successfully predict later concern for language development. While language measures were the most successful in predicting vocabulary in chapters 2 and 3, they did not continue to provide predictive power in chapter 4.

5.5 Summary of the discriminant function analyses

In each chapter, discriminant function analyses were used to establish the discriminatory ability of combinations of variables in distinguishing between children whose language was and was not typically developing. It is unlikely that there is one underlying cause of language delay and thus, a combination of variables should improve our ability to distinguish between these children (Conti-Ramsden & Durkin, 2012). In chapters 2 and 3, the success of classification accuracy was largely due to the power of earlier language measures. In chapter 3, when discriminating between typically developing and late talking children, the sensitivity of the models was poor when concurrent NWR was excluded. This was because the majority of children were not late talkers. In practice, this is a major issue; the majority of all children do not experience delays in vocabulary development and therefore, measures need to be sensitive enough to identify the few children who do experience delays among a larger number of children who do not experience delays.

Interestingly, in chapter 3, combining the variables together counter-intuitively reduced their ability to discriminate between typically developing and late talking children. Individually, earlier productive and receptive vocabulary, M3L and NWR could discriminate between these groups of children when using the 50 word cut off. Similarly, earlier gestures and NWR could discriminate between the groups using the failure to combine cut off. However, when combined, NWR contributed the most to classification accuracy, and the removal of NWR resulted in the loss of sensitivity. This shows that it is not that some measures are more accurate than others, in that some classify more children and others classify less children, but that different measures are classifying different subsets of children altogether. As a result, these measures cannot be used additively to improve classification accuracy. The implications of these results are that late talking children's performance differs across multiple measures. However, the finding that, when using the failure to combine classification, the sensitivity of the model was improved compared to the 50 word cut off, shows that there is some overlap in the performance of these late talking children across the measures.

These conflicting results from chapter 2 and chapter 3 suggest that individual variation and late talking may have different origins. It is possible that individual differences in the normal range are a result environmental differences, in that the same cognitive mechanisms support language acquisition in typically developing children but differences in environmental factors determine variance. However, where there is a diagnosable delay, it could be that there is a maturational difference in brain development. Therefore, lower typical development is not comparable to late talking. Thus, it is important to understand what is different between children with lower typically developing vocabulary skills and children who end up with a clinical diagnosis.

In chapter 4, the sensitivity of all of the discriminant function models was poor. Language skills between 15-18 months were no longer driving discrimination. This may be because the factors that cause difficulties in vocabulary acquisition early in life are different to those that cause difficulties in other (later acquired) aspects of language. As mentioned above, children who have language impairments later in childhood, have more difficulties with syntax (Bishop & Snowling, 2004; Leonard, 1998) or pragmatics (Bishop, 2014; Norbury & Bishop, 2002) than with vocabulary (though vocabulary is often also relatively low). Therefore, the factors underlying parental concern when children are older might be driven by different factors then those earlier in childhood.

5.6 Future studies

The research presented in this thesis has demonstrated the role of multiple measures in early vocabulary development and that some of these measures can be used to identify late talking children. In particular, early measures of vocabulary and gestures, syntax and NWR can be used to discriminate between children who are and are not experiencing a delay in vocabulary acquisition. However, the findings in chapter 4 emphasise that later concern for language development is not necessarily predicted by measures of vocabulary and gestures early in life. Later language impairment often presents as difficulties with syntax acquisition. Thus, syntax measures might be better suited for first predicting late talking and then for distinguishing between persisting and transient language impairment into later childhood. Therefore, future research should investigate whether or not syntax measures are a better predictor of later language learning in late talking children.

Further research should also focus on the relationship between speed of processing, NWR and syntax acquisition in late talking children. There is some research showing the relationship between speed of processing and syntax development (Peter et al., 2019; Fernald et al., 2006), however, no research to date looks at this relationship specifically in late talking children and children with language impairment. Similarly, the role of NWR and syntax has been previously explored (Adams & Gathercole, 2000; Blake, Austin, Cannon, Lisus, & Vaughan, 1991). Again, however, its role in predicting late talking as determined by a delay in syntax acquisition is unexplored.

Most importantly, future research needs to focus on finding predictors of language delay over time. In this research, the factors that could successfully distinguish between typically developing and late talking children were not the same factors that predicted concern for language development and an identified disability between 4-6 years. If there is a disconnect between vocabulary and syntax acquisition during language development, as predicted by some modularity theorists, this might explain why we find no relationship between early vocabulary skills and concern for language development between 4-6 years. Even if vocabulary and syntax are correlated early on in childhood, children's abilities could still vary as they get older. It is therefore important to establish, in terms of language delay, when this disconnect begins to emerge. If it is during early childhood, then it is important to examine syntax skills early on, and establish their role in language delay and late talking.

5.7 Conclusion

The three experimental chapters in this thesis have demonstrated that there are a variety of factors that predict variance in productive vocabulary at 24 and 25 months, and that many of these factors can be used to predict early language delay. However, early measures of language skills and risk factors cannot be used to predict if there will be concern for children's language development later in childhood. Chapters 2 and 3 replicated findings that early measures of vocabulary and gestures, mean length of utterance, speed of processing and NWR predict variance in productive vocabulary and expanded on these findings by demonstrating their unique role in vocabulary development. In addition, these chapters show the role of these factors in discriminating between children who do and do not have a delay in vocabulary acquisition. Chapter 4 demonstrated that even when combined, early measures of language, risk factors and parental concern cannot be used to discriminate between children who do and do not have concern for their language delay due to a lack of sensitivity. While the current thesis has provided important additional evidence for the unique role of multiple factors in early language acquisition and delay, more research is required to establish factors which can predict language impairment over time.

6 References

Acredolo, L., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development, 59*(2), 450–466. DOI: 10.2307/1130324.

Adams, A. M., & Gathercole, S. E. (2000). Limitations in working memory: implications for language development. International *Journal of Language & Communication Disorders, 35*(1), 95–116. https://doi.org/https://doi.org/10.1080/136828200247278

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development, 72*(2), 416–430. DOI: 10.1111/1467-8624.00287.

Alcock, K.J., Meints, K., Rowland, C. F., Christopher, A., Just, J. & Brelsford, J (2020). The UK communicative development inventory: Words and gestures.

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics, 2(*3), 370-374. DOI: 10.1002/wics.84.

All Party Parliamentary Group on Speech and Language Difficulties. (2013). The links between speech, language and communication needs and social disadvantage. Retrieved from https://www.rcslt.org/governments/docs/all_party_parliamentary_group_on_slcn_inquiry_report

Barre, N., Morgan, A., Doyle, L. W., & Anderson, P. J. (2011). Language abilities in children who were very preterm and/or very low birth weight: A meta-analysis. *The Journal of Pediatrics, 158*(5), 766–774. https://doi.org/10.1016/j.jpeds.2010.10.032

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. DOI: 10.1007/0-387-22747-4_4

Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language and Cognitive Processes, 12*(5/6), 507–584. https://doi.org/10.1080/016909697386628

Bates, E., & Goodman, J. C. (1999). On the emergence of grammar from the lexicon. On the emergence of grammar from the lexicon. In B. MacWhinney (ed.), *The emergence of language*. Mahwah, NJ: Erlbaum.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., … Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language, 21*, 85–123. https://doi.org/10.1017/S0305000900008680

Bidgood, A., Durrant , S., Peter, M., Pine, J. M., & Rowland, C.F., (unpub.). The Language 0-5 Project non-word repetition test. Unpublished manuscript, University of Liverpool. Available from https://osf.io/kau5f/. DOI: 10.17605/OSF.IO/KAU5F.

Bishop, D. V. M. (2014). Pragmatic language impairment: A correlate of SLI, a distinct subgroup, or part of the autistic continuum? In speech and language impairments in children: Causes, characteristics, intervention and outcome (pp. 113–128). *Psychology Press.* https://doi.org/10.4324/9781315784878

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry, 31*(7), 1027–1050. DOI: 10.1111/j.1469-7610.1990.tb00844.x.

Bishop, D. V. M., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of Child Psychology and Psychiatry, 37*(4), 391–403. DOI: 10.1111/j.1469-7610.1996.tb01420.x.

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52*, 156–173.

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same of different? *Psychological Bulletin, 160*(6), 858–886. https://doi.org/10.1037/0033-2909.130.6.858

Bishop, D. V. M., Holt, G., Line, E., McDonald, D., McDonald, S., & Watt, H. (2012). Parental phonological memory contributes to prediction of outcome of late talkers from 20 months to 4 years: A longitudinal study of precursors of specific language impairment. *Journal of Neurodevelopmental Disorders, 4*(1), 1–12. https://doi.org/https://doi.org/10.1186/1866-1955-4-3

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Adams, C., Archibald, L., … Whitehouse, A. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. PLoS ONE, 11(7), 1–26. https://doi.org/10.1371/journal.pone.0158753

Blake, J., Austin, W., Cannon, M., Lisus, A., & Vaughan, A. (1991). The relationship between memory span and measures of imitative and spontaneous language complexity in preschool children. *International Journal of Behavioural Development, 17*(1), 91–107. https://doi.org/https://doi.org/10.1177/016502549401700106

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Lexical leverage: category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental Science, 19*(6), 918–932. DOI: 10.1111/desc.12343.

Botting, N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language & Communication Disorders, 36*(4), 421-432.

Bright Futures Steering Committee & Medical Home Initiatives for Children with Special Needs Project Advisory Committee. (2006). Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *American Academy of Pediatrics, 118*(1), 405–420. DOI: 10.1542/peds.2006-1231.

Brown, R. (1973). Brown, R. (1973). *A first language: The early stages*. MA: Harvard U. Press.

Campbell, T. F., Dollaghan, C. A., Rockette, H. E., Paradise, J. L., Feldman, H. M., Shriberg, L. D., … Kurs-Lasky, M. (2003). Risk factors for speech delay of unknown origin in 3-year-old children. *Child Development, 74*(2), 346–357. https://doi.org/10.1111/1467-8624.7402002

Carroll, J. M., & Breadmore, H. L. (2017). Not all phonological awareness deficits are created equal: evidence from a comparison between children with Otitis Media and poor readers. *Developmental Science, 21*(3), 1–12. https://doi.org/10.1111/desc.12588

Carson, D. K., & Gavin, W. J. (1998). Concurrent and predictive validity of an early language screening program family supportive milieu: A study of protective factors and mechanisms among the invulnerable adolescents view project. *Journal of Speech Language and Hearing Research, 41,* 627–641. https://doi.org/10.1044/jslhr.4103.627

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.

Cherry, L., & Lewis, M. (1976). Mothers and two-year-olds: A study of sex-differentiated aspects of verbal interaction. *Developmental Psychology, 12*(4), 278-282. DOI: 10.1002/j.2333-8504.1975.tb01077.x.

Chiat, S., & Roy, P. (2008). Early phonological and sociocognitive skills as predictors of later language and social communication outcomes. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 49*(6), 635–645. https://doi.org/10.1111/j.1469-7610.2008.01881.x

Collisson, B. A., Graham, S. A., Preston, J. L., Rose, M. S., McDonald, S., & Tough, S. (2016). Risk and protective factors for late talking: An epidemiologic investigation. *Journal of Pediatrics, 172*, 168-174.e1. https://doi.org/10.1016/j.jpeds.2016.02.020

Colonnesi, C., Stams, G. J. J. M., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review, 30*(4), 352–366. DOI: 10.1016/j.dr.2010.10.001.

Conti-Ramsden, G., & Durkin, K. (2012). Language development and assessment in the preschool period. *Neuropsychol Rev, 22*, 384–401. https://doi.org/10.1007/s11065-012-9208-z

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry, 42*(6), 741–748.

Curtin, S., & Werker, J. F. (2007). The perceptual foundations of phonological development. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 579–599). Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780198568971.013.0035.

Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay I: Predicting persistent and transient delay at 3 and 4 years. *Journal of Speech, Language and Hearing Research, 46*, 544-560. DOI: 10.1044/1092-4388(2003/044).

Demir-Vegter, S., Aarts, R., & Kurvers, J. (2014). Lexical richness in maternal input and vocabulary development of Turkish preschoolers in the Netherlands. *J Psycholinguist Res, 43*, 149–165. https://doi.org/10.1007/s10936-013-9245-7

Department for Work and Pensions. (2019). *Households below average income: An analysis of the UK income distribution: 1994/95-2014/15*. Retrieved from https://www.gov.uk/government/statistics/households-below-average-income-199495-to-201415.

Department of Health. (2009). *Healthy Child Programme: The two year review*. Retrieved from https://www.gov.uk/government/publications/healthy-child-programme-review-of-children-aged-2

Dimitrova, N., Özçalışkan, Ş., & Adamson, L. B. (2016). Parents' translations of child gesture facilitate word learning in children with autism, Down syndrome and typical development. *Journal of Autism and Developmental Disorders, 46*(1), 221-231. DOI: 10.1007/s10803-015-2566-7.

Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child development, 74*(2), 394-412. DOI: 10.1111/1467-8624.7402005.

Dixon, J. A., & Marchman, V. A. (2007). Grammar and the lexicon: Developmental ordering in language acquisition. *Child Development, 78*(1), 190–212. https://doi.org/10.1111/j.1467-8624.2007.00992.x

Duff, F. J., Nation, K., Plunkett, K., & Bishop, D. V. M. (2015). Early prediction of language and literacy problems: Is 18 months too early? *PeerJ, 3*, 1-12. DOI: 10.7717/peerj.1098.

Eaton, C. T., Newman, R. S., Rathner, R. N., & Rowe, M. L. (2015). Non-word repetition in 2-year-olds: Replication of an adapted paradigm and a useful methodological extension. *Clinical Linguistics and Phonetics, 29(*7), 523–535.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology, 49*(1), 234–250.

Evans, G. W., Maxwell, L. E., & Hart, B. (1999). Parental language and verbal responsiveness to children in crowded homes. *Developmental Psychology, 35*(4), 1020–1023. https://doi.org/10.1037/0012-1649.35.4.1020

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in Child Development, 59*(5), 174–185.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E (2007). *MacArthur-Bates communicative development inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Brookes.

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically-developing and late-talking toddlers. *Child Development, 83*(1), 203–222.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology, 42*(1), 98–116.

Fischel, J. E., Whitehurst, G. J., Caulfield, M. B., & Debaryshe, B. (1989). Language growth in children with expressive language delay. *Pediatrics, 83*, 218–227.

Gathercole, S. E., & Adams, A. M. (1993). Phonological working memory in very young children. *Developmental Psychology*, 29(4), 770–778. DOI: 10.1037/0012-1649.29.4.770.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200–213.

Gathercole, S. E., & Baddeley, A. D. (1990a). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*(3), 336–360.

Gathercole, S. E., & Baddeley, A. D. (1990b). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*(4), 439–454.

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Kimbrough Oller, D., ... & Paul, T. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology, 26*(2), 248-265. https://doi.org/10.1044/2016_ajslp-15-0169.

Girolametto, L., Wiigs, M., Smyth, R., Weitzman, E., & Pearce, P. (2001). Children with a history of expressive vocabulary delay: Outcome at 5 years of age.

American Journal of Speech-Language Pathology, 10(4), 358–369. https://doi.org/10.1044/1058-0360(2001/030)

Glascoe, F. P. (1991). Can clinical judgment detect children with speech-language problems? *Pediatrics, 87*(3), 317–322. Retrieved from www.aappublications.org/news

Glascoe, F. P., & Dworkin, P. H. (1995). The role of parents in the detection of developmental and behavioural problems. *Pediatrics, 95*(6), 829–836.

Glascoe, F. P., Altemeier, W. A., & MacLean, W. E. (1989). The importance of parents' concerns about their child's development. *American Journal of Diseases of Children, 143*(8), 955–958. https://doi.org/10.1001/archpedi.1989.02150200115029

Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: rate, content, and the vocabulary spurt. *Journal of Child Language, 17*, 171–183. https://doi.org/10.1017/S0305000900013167

Goldin-Meadow, S., Goodrich, W., Sauer, E., & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental science, 10*(6), 778-785. DOI: 10.1111/j.1467-7687.2007.00636.x.

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Development, 90*(3), 985–992. https://doi.org/10.1111/cdev.13128

Guarini, A., Sansavini, A., Fabbri, C., Alessandroni, R., Faldella, G., & Karmiloff-Smith, A. (2009). Reconsidering the impact of preterm birth on language outcome. *Early Human Development, 85*, 639–645. https://doi.org/10.1016/j.earlhumdev.2009.08.061

Hadley, P. A., & Holt, J. K. (2006). Individual differences in the onset of tense marking: A growth-curve analysis. *Journal of Speech, Language, and Hearing Research, 49*, 984-1000. DOI: 10.1044/1092-4388(2006/071).

Hadley, P. A., & Short, H. (2005). The onset of tense marking in children at risk for specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*, 1344–1362. https://doi.org/10.1044/1092-4388(2005/094)

Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine, 4*(2), 627.

Harrison, L. J., & McLeod, S. (2010). Risk and protective factors associated with speech and language impairment in a nationally representative sample of 4-to 5-year-old children. *Journal of Speech, Language, and Hearing Research, 53*, 508–529. https://doi.org/10.1044/1092-4388(2009/08-0086

Hart, B., & Risley, T. R. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology, 28*(6), 1096-1105. DOI: 10.1037/0012-1649.28.6.1096.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in everyday experience of young American children*. Baltimore, MD: Paul F Brookes.

Henrichs, J., Rescorla, L., Schenk, J. J., Schmidt, H. G. G., Jaddoe, V. W. V., Hofman, A., … Tiemeier, H. (2011). Examining continuity of early expressive vocabulary development: the generation R study. *Journal of Speech Language and Hearing Research, 54*(3), 854–869. https://doi.org/10.1044/1092-4388(2010/09-0255)854

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., … Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science, 26*(7), 1071–1083. DOI: 10.1177/0956797615581493.

Hoff-Ginsberg, E. (1991). Mother-child conversation in different social classes and communicative settings. *Child development, 62*(4), 782-796. DOI: 10.2307/1131177.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child development, 74*(5), 1368-1378. DOI: 10.1111/1467-8624.00612.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development Stable, 73*(2), 418–433. Retrieved from ttp://www.jstor.org/stable/3696366

Hoff, E., Core, C., & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20- to 24-month-olds. *Journal of Child Language, 35*(4), 903–916.

Horwitz, S. M., Irwin, J. R., Briggs-Gowan, M. J., Bosson Heenan, J. M., Mendoza, J., Carter, A. S., … Carter, A. S. (2003). Language delay in a community cohort of young children. *Journal of the American Academy of Child &*

Adolescent Psychiatry, 42(8), 932–940. https://doi.org/10.1097/01.CHI.0000046889.27264.5E

Hsu, H. C., & Iyer, S. N. (2016). Early gesture, early vocabulary, and risk of language impairment in preschoolers. *Research in Developmental Disabilities, 57*, 201–210. DOI: 10.1016/j.ridd.2016.06.012.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science, 11*(6), 1–16. https://doi.org/10.1111/j.1467-7687.2008.00768.x.Does

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*(2), 236–248. https://doi.org/10.1037/0012-1649.27.2.236

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology, 61*(4), 343-365. DOI: 10.1016/j.cogpsych.2010.08.002.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*(5), 367–371. DOI: 10.1111/j.0956-7976.2005.01542.x.

Jackman, S. (2010). pscl: Classes and methods for R. Developed in the Political Science Computational Laboratory, Stanford University. Department of Political Science, Stanford University, Stanford, CA. R package version 1.03. 5. Retrieved from http://www. pscl. stanford. edu/.

Jansson-Verkasalo, E., Valkama, M., Vainionpää, L., Pääkkö, E., Ilkko, E., & Lehtihalmes, M. (2004). Language development in very low birth weight preterm children: A follow-up study. *Folia Phoniatrica et Logopaedica, 56*, 108–119.

Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive Psychology, 98*, 1–21. DOI: 10.31234/osf.io/wdbpc.

Jones, G., Gobet, F., & Pine, J. M. (2007). Linking working memory and long-term memory: a computational model of the learning of new words. *Developmental Science, 10*(6), 853-873. DOI: 10.1111/j.1467-7687.2007.00638.x.

Jordan, J. A., & Coulter, L. (2017). Individual differences in speech and language ability profiles in areas of high deprivation. *Child Care in Practice, 23*(4), 417–430. https://doi.org/10.1080/13575279.2016.1188759

Kennedy, C. R., McCann, D. C., Campbell, M. J., Law, C. M., Mullee, M., Petrou, S., … Stevenson, J. (2006). Language ability after early detection of permanent childhood hearing impairment. *New England Journal of Medicine, 354*(20), 2131–2141. https://doi.org/10.1056/NEJMoa054915

Kidd, E., & Donnelly, S. (2019). Individual differences in first language acquisition. *Annual Review of Linguistics,6,* 1–22. https://doi.org/10.1146/annurev-linguistics-011619-030326

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences, 22*(2), 154–169. doi.org/10.1016/j.tics.2017.11.006

Killing, S. E., & Bishop, D. V. M. (2008). Move it! Visual feedback enhances validity of preferential looking as a measure of individual differences in vocabulary in toddlers. *Developmental Science, 11*(4), 525–530. DOI: 10.1111/j.1467-7687.2008.00698.x.

Klee, T., Carson, D. K., Gavin, W. J., Hall, H., Kent, A., & Reece, S. (1998). Concurrent and predictive validity of an early language screening program. *Journal of Speech Language and Hearing Research, 41*, 627–641. DOI: 10.1044/jslhr.4103.627.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. *International Journal of Language & Communication Disorders, 35*, 165–188. DOI: 10.1080/136828200247133.

Law, J., McBean, K., & Rush, R. (2011). Communication skills in a population of primary school-aged children raised in an area of pronounced social disadvantage. *International Journal of Language and Communication Disorders, 46*(6), 657–664. https://doi.org/10.1111/j.1460-6984.2011.00036.x

LENA Research Foundation. (2014). The LENA Research Foundation. Retrieved from http://www.lenafoundation.org/

Leonard, L. B. (1998). *Children with Specific Language Impairment.* MA: The MIT Press. https://doi.org/10.7551/mitpress/9152.001.0001

Leonard, L. B. (2014). *Children with specific language impairment* (2nd Edition). Cambridge, MA: MIT Press.

Locke, J. L. (1997). A theory of neurolinguistic development. *Brain and Language, 58*, 265–326. https://doi.org/10.1006/brln.1997.1791

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. (3rd Edition). Mahwah, NJ: Lawrence Erlbaum Associates.

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: a test of the critical mass hypothesis. *Journal of Child Language, 21*(2), 339–366. https://doi.org/10.1017/S0305000900009302

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood: Fast-Track Report. *Developmental Science, 11*(3), 9–16. https://doi.org/10.1111/j.1467-7687.2008.00671.x

Marchman, V. A., Adams, K. A., Loi, E. C., Fernald, A., & Feldman, H. M. (2016). Early language processing efficiency predicts later receptive vocabulary outcomes in children born preterm. *Child Neuropsychology,* 22(6), 649-665. DOI: 10.1080/09297049.2015.1038987.

McGregor, K., Sheng, L., & Smith, B. (2005). The precocious two-year-old: Status of the lexicon and links to the grammar. *Journal of Child Language, 32*, 563–585. DOI: 10.1017/s0305000905006926.

Meints, K, Fletcher, K & Just, J. (2017) Lincoln Toddler Communicative Development Inventory. https://cpb-eu-w2.wpmucdn.com/blogs.lincoln.ac.uk/dist/b/6736/files/2017/11/Lincoln_toddler_cdiv2-2.pdf

Mintz, T. H., & Gleitman, L. R. (2002). Adjectives really do modify nouns: The incremental and restricted nature of early adjective acquisition. *Cognition, 84*(3), 267–293. https://doi.org/10.1016/S0010-0277(02)00047-1

Mirman, D. (2014). Growth curve analysis and visualization using R. CRC Press. DOI: 10.1201/9781315373218.

Moyle, M. J., Weismer, S. E., Evans, J. L., & Lindstrom, M. J. (2007). Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *Journal of Speech, Language, and Hearing Research, 50*(8), 508–528. https://doi.org/10.1044/1092-4388(2007/035)

Nelson, H. D., Nygren, P., Walker, M., & Panoscha, R. (2006). screening for speech and language delay in preschool children: systematic evidence review for the US Preventive Services Task Force. *Pediatrics, 117*(2), e298–e319. https://doi.org/10.1542/peds.2005-1467

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38*, 1–135. DOI: 10.2307/1165788.

Norbury, C. F., & Bishop, D. V. M. (2002). Inferential processing and story recall in children with communication problems: A comparison of specific language impairment, pragmatic language impairment and high-functioning autism. *International Journal of Language and Communication Disorders, 37*(3), 227–251. https://doi.org/10.1080/13682820210136269

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Olswang, L. B., Rodriguez, B., & Timler, G. (1998). Recommending intervention for toddlers with specific language learning difficulties: We may not have all the answers, but we know a lot. *American Journal of Speech-Language Pathology, 7*(1), 23–32. DOI: 10.1044/1058-0360.0701.23.

Özçalişkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition, 96*(3), 101–113. DOI: 10.1016/j.cognition.2005.01.001.

Özçalışkan, Ş., Adamson, L. B., Dimitrova, N., & Baumann, S. (2017). Early gesture provides a helping hand to spoken vocabulary development for children with autism, Down syndrome, and typical development. *Journal of Cognition and Development, 18*(3), 325-337. DOI: 10.1080/15248372.2017.1329735.

Paul, R. (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech-Language Pathology, 5*, 5–30.

Paul, R. (2000). Predicting outcomes of early expressive language delay: Ethical implications. In D. V. M. Bishop & L. B. Leonard (Eds.) *Speech and* *language impairments in children: Causes, characteristics, intervention and outcome* (pp. 195–209). Hove: Psychology Press.

Paul, R., & Alforde, S. (1993). Grammatical morpheme acquisition in 4-year-olds with normal, impaired, and late-developing language. *Journal of Speech, Language, and Hearing Research 36*(6), 1271–1275. https://doi.org/10.1044/jshr.3606.1271

Paul, R., Looney, S. S., & Dahm, P. S. (1991). Communication and socialization skills at ages 2 and 3 in "late-talking" young children. *Journal of Speech, Language, and Hearing Research, 34*(4), 858–865.

Paul, R., & Roth, F. P. (2011). Characterizing and predicting outcomes of communication delays in infants and toddlers: Implication for clinical practice. *Language Speech and Hearing Services in School, 42*, 331–340. doi.org/10.1044/0161-1461(2010/09-0067)b

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics, 18*(1). 41–58. DOI: 10.1017/s0142716400009863.

Peter, M. S., Durrant, S., Jessop, A., Bidgood, A., Pine, J. M., & Rowland, C. F. (2019). Does speed of processing or vocabulary size predict later language growth in toddlers? *Cognitive Psychology, 115*,1-25. DOI: 10.1016/j.cogpsych.2019.101238

Petinou, K., & Spanoudis, G. (2014). Early language delay phenotypes and correlation with later linguistic abilities. *Folia Phoniatrica et Logopaedica, 66*, 67–76. DOI: 10.1159/000365848.

Quinn, S., Donnelly, S., & Kidd, E. (2018). The relationship between symbolic play and language acquisition: A meta-analytic review. *Developmental Review, 49*, 121–135. DOI: 10.1016/j.dr.2018.05.005.

Ramer, A. L. H. (1976). Syntactic styles in emerging language. *Journal of Child Language, 3*(1), 49–62. https://doi.org/10.1017/S0305000900001306

Reilly, S., Bavin, E. L., Bretherton, L., Conway, L., Eadie, P., Cini, E., … Wake, M. (2009). The Early Language in Victoria Study (ELVS): A prospective, longitudinal study of communication skills and expressive vocabulary development at 8, 12 and 24 months. *International Journal of Speech-Language Pathology, 11*(5), 344–357. DOI: 10.1080/17549500903147560.

Reilly, S., Wake, M., Bavin, E. L., Prior, M., Williams, J., Bretherton, L., … Ukoumunne, O. C. (2007). Predicting language at 2 years of age: A prospective community study. *Pediatrics, 120*(6), 1441–1449. DOI: 10.1542/peds.2007-0045.

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., … Bretherton, L. (2010). Predicting language outcomes at 4 years of age: findings from Early Language in Victoria Study. *Pediatrics, 126*(6), 1530–1537. DOI: 10.1542/peds.2010-0254.

Rescorla, L. (1989). The Language Development Survey: A screening for delayed language in toddlers. *Journal of Speech and Hearing Disorders, 54,* 587–599. DOI: 10.1044/jshd.5404.587.

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 45*(2), 360–371. DOI: 10.1044/1092-4388(2002/028).

Rescorla, L. (2005). Age 13 language and reading outcomes in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 48*(2), 459-472. DOI: 10.1044/1092-4388(2005/031).

Rescorla, L. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language, and Hearing Research, 52*(1), 16-30. DOI: 10.1044/1092-4388(2008/07-0171).

Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews, 17*(2), 141–150. DOI: 10.1002/ddrr.1108.

Rescorla, L., & Achenbach, T. M. (2002). Use of the Language Development Survey (LDS) in a national probability sample of children 18 to 35 months old. *Journal of Speech, Language, and Hearing Research, 45*(4), 733–743. https://doi.org/10.1044/1092-4388(2002/059)

Rescorla, L., Dahlsgaard, K., & Roberts, J. (2000). Late-talking toddlers: MLU and IPSyn outcomes at 3; 0 and 4; 0. *Journal of Child Language, 27*(3), 643-664. DOI: 10.1017/s0305000900004232

Rescorla, L., & Schwartz, E. (1990). Outcome of toddlers with specific expressive language delay. *Applied Psycholinguistics, 11*(04), 393. doi.org/10.1017/s0142716400009644

Rice, M. L., Taylor, C. L., & Zubrick, S. R. (2008). Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *Journal of Speech, Language, and Hearing Research, 51*(2), 394–407. https://doi.org/10.1044/1092-4388(2008/029)

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. E. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science, 29*(5), 700–710. DOI: 10.1177/0956797617742725.

Roos, E. M., & Ellis Weismer, S. (2008). Language outcomes of late talking toddlers at preschool and beyond. *Perspectives on Language Learning and Education, 15*(3), 119–126. https://doi.org/10.1044/lle15.3.119

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83*(5), 1762–1774.

Rowe, M. L. (2018). Understanding socioeconomic differences in parents' speech to children. *Child Development Perspectives, 12*(2), 122–127. https://doi.org/10.1111/cdep.12271

Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science, 12*(1), 182–187. DOI: 10.1111/j.1467-7687.2008.00764.x.

Rowe, M. L., Özçalişkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language, 28*(2), 182–199. DOI: 10.1177/0142723707088310.

Rowe, M. L., Pan, B. A., & Ayoub, C. (2009). Predictors of variation in maternal talk to children: A longitudinal study of low-income families. *Parenting: Science and Practice, 5*(3), 285–310. https://doi.org/10.1207/s15327922par0503_3

Rowland, C.F., Bidgood, A., Durrant, S., Peter, M., & Pine, J. M. (unpub.). The Language 0-5 Project. Unpublished manuscript, University of Liverpool. Available from https://osf.io/kau5f/. DOI: 0.17605/OSF.IO/KAU5F

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2- to 4- year-olds: Evidence from typically developing children.

Journal of Speech, Language, and Hearing Research, 47(1), 223-234. DOI: 10.1044/1092-4388(2004/019).

Save the Children. (2015). *Ready to read: Closing the gap in early language skills so that every child in England can read well*. Retrieved from https://www.savethechildren.org.uk/content/dam/global/reports/education-and-child-protection/ready-to-read.pdf

Schachter, F. F., Shore, E., Hodapp, R., Chalfin, S., & Bundy, C. (1978). Do girls talk earlier?: Mean length of utterance in toddlers. *Developmental Psychology, 14*(4), 388–392. https://doi.org/10.1037/0012-1649.14.4.388

Schults, A., Tulviste, T., & Konstabel, K. (2012). Early vocabulary and gestures in Estonian children. *Journal of Child Language, 39*(3), 664–686. DOI: 10.1017/s0305000911000225.

Singleton, C. N. (2018). Late talkers: Why the wait-and-see approach is outdated. *Pediatric Clinics of North America, 65*, 13–29. https://doi.org/10.1016/j.pcl.2017.08.018

Smith, A. B., & Jackins, M. (2014). Relationship between longest utterances and later MLU in late talkers. *Clinical Linguistics & Phonetics, 28*(3), 143–152. DOI: 10.3109/02699206.2013.839746.

Stanton-Chapman, T. L., Chapman, D. A., Bainbridge, N. L., & Scott, K. G. (2002). Identification of early risk factors for language impairment. *Research in Developmental Disabilities, 23*(390–405). https://doi.org/10.1016/s0891-4222(02)00141-5

Stokes, S. F., & Klee, T. (2009). The diagnostic accuracy of a new test of early nonword repetition for differentiating late talking and typically developing children. *Journal of Speech and Hearing Research, 52*, 872–883. DOI: 10.1044/1092-4388(2009/08-0030).

Stolt, S., Haataja, L., Lapinleimu, H., & Lehtonen, L. (2009). The early lexical development and its predictive value to language skills at 2 years in very-low-birth-weight children. *Journal of Communication Disorder, 42,* 107–123. https://doi.org/10.1016/j.jcomdis.2008.10.002

Stolt, S., Lind, A., Matomäki, J., Haataja, L., Lapinleimu, H., & Lehtonen, L. (2016). Do the early development of gestures and receptive and expressive language predict language skills at 5;0 in prematurely born very-low-birth-weight children? *Journal of Communication Disorders, 61*, 16–28. https://doi.org/10.1016/j.jcomdis.2016.03.002

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. Journal of Phonetics, 39(4), 456-466. DOI: 10.1016/j.wocn.2010.09.001.

Sylva, K., Stein, A., Leach, P., Barnes, J., & Malmberg, L. E. (2011). Effects of early child-care on cognition, language, and task-related behaviours at 18 months: An English study. British *Journal of Developmental Psychology, 29*(1), 18–45. https://doi.org/10.1348/026151010X533229

Thal, D. J., Bates, E., Goodman, J., & Jahn-Samilo, J. (1997). Continuity of language abilities: An exploratory study of late-and early-talking toddlers. *Developmental Neuropsychology, 13*(3), 239–273. DOI: 10.1080/87565649709540681.

Thal, D. J., Miller, S., Carlson, J., & Vega, M. M. (2005). Nonword repetition and language development in 4-year-old children with and without a history of early language delay. *Journal of Speech, Language and Hearing Research, 48*(6), 1481–1495. DOI: 10.1044/1092-4388(2005/103).

Thal, D. J., Tobias, S., & Morrison, D. (1991). Language and gesture in late talkers: A 1-year followup. *Journal of Speech and Hearing Research, 34*, 604–612.DOI: 10.1044/jshr.3403.604.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of speech, language, and hearing research, 40*(6), 1245-1260.

Torrington Eaton, C., Newman, R. S., Ratner, N. B., & Rowe, M. L. (2015). Non-word repetition in 2-year-olds: Replication of an adapted paradigm and a useful methodological extension. *Clinical Linguistics & Phonetics, 29*(7), 523-535. DOI: 10.3109/02699206.2015.1029594.

Turner, I. F. (1974). Cornitive effects of playgroup attendance. *The Irish Journal of Education, 8*(1), 30–35.

Valian, V. (2014). Arguing about innateness. *Journal of child language*, *41*(S1), 78-92.

Van Der Lely, H. K. J., Rosen, S., & McClelland, A. (1998). Evidence for a grammar-specific deficit in children. *Current Biology, 8*(23), 1253–1258. https://doi.org/10.1016/S0960-9822(07)00534-9

Van Noort-Van Der Spek, I. L., Franken, M.-C. J. P., & Weisglas-Kuperus, N. (2012). Language functions in preterm-born children: A systematic review and meta-analysis. *Pediatrics, 129*(4), 745–754. https://doi.org/10.1542/peds.2011-1728

Vernon-Feagans, L., Garrett-Peters, P., Willoughby, M., Mills-Koonce, R., Cox, M., Blair, C., … Willoughby, M. (2012). Chaos, poverty, and parenting: Predictors of early language development. *Early Childhood Research Quarterly, 27*(3), 339–351. https://doi.org/10.1016/j.ecresq.2011.11.001

Vigil, D. C., Hodges, J., & Klee, T. (2005). Quantity and quality of parental language input to late-talking toddlers during play. *Child Language Teaching and Therapy, 21*(1), 107–122. DOI: 10.1191/0265659005ct284oa.

Warnes, G. R., Bolker, B., Lumley, T. & Johnson, R. C. (2018). Package ‘gmodels’.

Watt, N., Wetherby, A., & Shumway, S. (2006). Prelinguistic predictors of language outcome at 3 years of age. *Journal of Speech Language and Hearing Research, 49*, 1224-1237. DOI: 10.1044/1092-4388(2006/088).

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*(11), 2143–2152. DOI: 10.1177/0956797613488145.

Weismer, S. E., Murray-Branch, J., & Miller, J. F. (1994). A prospective longitudinal study of language development in late talkers. *Journal of Speech & Hearing Research, 37*(4), 852–867. https://doi.org/10.1044/jshr.3704.852

Weismer, S. E., Venker, C. E., Evans, J. L., & Moyle, M. J. (2014). Fast mapping in late-talking toddlers. *Applied Psycholinguistics, 34*(1), 69-89. https://doi.org/10.1017/S0142716411000610

Westerlund, M., Berglund, E., & Eriksson, M. (2006). Can severely language delayed 3-year-olds be identified at 18 months? Evaluation of a screening version of the MacArthur-Bates Communicative Development Inventories. *Journal of Speech, Language and Hearing Research, 49*(2), 237–247. DOI: 10.1044/1092-4388(2006/020).

Whitehurst, G. J., & Fischel, J. E. (1994). Practitioner review: Early developmental language delay: What, if anything, should the clinician do about it? *Journal of Child Psychology and Psychiatry, 35*(4), 613–648. https://doi.org/10.1111/j.1469-7610.1994.tb01210.x

Winskel, H. (2006). The effects of an early history of otitis media on children's language and literacy skill development. *British Journal of Educational Psychology, 76*(4), 727–744. https://doi.org/10.1348/000709905X68312

Wordbank (2016). Wordbank: an open database of children's vocabulary development. Retrieved from http://wordbank.stanford.edu/.

Zambrana, I. M., Ystrom, E., Schjølberg, S., & Pons, F. (2013). Action imitation at 1½ years is better than pointing gesture in predicting late development of language production at 3 years of age. *Child Development, 84*(2), 560–573. DOI: 10.1111/j.1467-8624.2012.01872.x.

Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., & Yapanel, U. (2009). Teaching by listening: The importance of adult-child conversations to Language Development. *Pediatrics, 124*(1), 342–349. DOI: 10.1542/peds.2008-2267.

Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. W. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech Language and Hearing Research, 50*(6), 1562–1592. DOI: 10.1044/1092-4388(2007/106).