The Economic Effects of Childhood Speech and Language Difficulties: A Cross-Cohort Quantitative Life Course Analysis

by

Joseph Willoughby

School of Education, Communication and Language Sciences

Newcastle University

September 2020

A thesis submitted to the School of Education, Communication and Language Sciences of Newcastle University for the degree of Doctor of Philosophy.

Abstract

Introduction

Early vocabulary and speech development predict poorer broader childhood skills, educational attainment, mental health, and employment. However, little is known beyond early adulthood about how language relates to income, non-cognitive skills, or physical health.

Aims

The thesis aims to utilise historical data to highlight the broad and enduring impact of early communication over the life course. Appropriate, evidence-based recommendations are established for research and clinical practice.

Methodology

Two datasets were selected, capable of modelling the importance of early communication skills on life chances. The Newcastle Thousand Families Study (1947) incorporates novel health visitor observation data from age two to seven of 881 children, identifying 161 with Speech, Language and Communication Needs. The British Cohort Study (1970) complements this analysis with a national dataset of thousands of children, containing a continuous assessment of receptive vocabulary at age five. Regression, mediation, and counter-factual econometric methods (Oaxaca-Blinder Decomposition and Regression Discontinuity Design) are utilised to maximise the value in the historical data.

Findings

SLCN is associated with introverted personality traits, and a reduced tendency to enjoy new experiences in later adulthood. Lower earnings and health were observed at age 50 but not age 62.

Mediation analysis confirms the existence of a direct relationship between early receptive vocabulary and later earnings across adulthood. No significant mediation effect was observed through education or non-cognitive skills after accounting for this direct effect.

However, when comparing income differences between high and low vocabulary groups (no direct effect of vocabulary), higher levels of educational attainment and cognitive skills were found in the high vocabulary group on average, explaining most of the observed group

differences in income. Returns to these variables are not statistically different between vocabulary groups, implying that vocabulary does not moderate the importance of known drivers of income. In summary, it is disadvantage in early years rather than discrimination in adulthood which explains the income gap.

Furthermore, whilst vocabulary scores predict a range of functional outcomes, this is not true over narrow thresholds. Small differences in language scores do not predict outcomes, regardless of the placement of the threshold, suggesting that thresholds are an unreliable method of identifying functional language disorders. Therefore, strictly applied thresholds should be removed from clinical diagnoses, and academic research should reduce its reliance on unsupported distinctions.

Conclusion

Varying methodological approaches and data enable the illustration of substantial positive associations between early verbal communication skills and a range of previously unmeasured lifelong outcomes.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I am enormously grateful to my supervisors for their patience and guidance. To Professor James Law for your engaging discussions and expertise. Our conversations and your feedback guided me in developing my understanding of the broad field of speech and language development. Particularly, your comments on my writing style helped me to focus every chapter and to maintain a consistent narrative. I am thankful for your encouragement and for involving me in many different research communities. To Professor John Wildman, for providing balance between specific econometric advice and a constant appreciation of the big picture. Your feedback was always constructive, efficiently communicated, and valuable. Both of my supervisors modelled the attitudes that I intend to follow throughout my career.

The Economic and Social Research Council deserve acknowledgement for their funding, without which this research would not have happened. Their support of multidisciplinary research continues to support social scientists to make a difference.

I would also like to thank all the participants and managers of the datasets used within this thesis. Thousands of individuals taking time to respond to lengthy questionnaires over decades enables all the analysis presented here and in many other publications. The Centre for Longitudinal Studies manage the British Cohort Study 1970 used in Chapters 3 to 5 and provide informative documentation of the many measures used. Within the Newcastle Thousand Families Study team, I received a huge amount of personal support and inspiration. In particular, I am grateful to Katherine Kirton for digitising 881 series of health visitor records to allow the development of speech and language profiles and to Dr Kay Mann for providing cleaned data in Stata for many of the variables used in the analysis in Chapter 2. The whole team at the Sir James Spence Institute help to maintain this resource and provide a welcoming environment for new researchers.

I also must thank a number of colleagues who made coming to work enjoyable and rewarding throughout my years in Newcastle. Dr Robert Anderson, Dr Helen Stringer, Dr Helen Wareham, Professor Ghada Khattab and Professor Cristina McKean all provided me with opportunities that expanded my horizons through teaching, presenting and research opportunities. Their trust in PhD students made the whole experience fulfilling and inspiring.

I was lucky to be part of a fantastic work environment in the G16 office and beyond. I had the best shared office with Josie, Anastasia, Sarah and Helen. The Child Language Research Group further allowed me to develop my ideas and presentation skills within a friendly and informative atmosphere. The economics PhD society, “NERD”, for allowing me to present at the earliest stages of ideas and for guaranteeing a beer would be on the cards after presenting. In particular, to Scott Kirkman for running this society, helping me to code in stata when I reached my limit, and for being a top mate, housemate and later spare bed provider.

This thesis would not have happened without my family behind me. I thank my mum and dad for constant support and allowing me to live at home when the funding expired. James and Sophie for constantly asking “so what is it you do?” to keep me communicative. And to Harry for always having time to listen to my work and for being a perfect antidote to stress.

Declaration

I declare that the thesis presented is original work, produced under the supervision of Professors James Law and John Wildman. The initial ideas behind each analysis were discussed in meetings with both supervisors. Feedback was then provided on the work drafted from these discussions, improving the clarity and accuracy of the work presented. They provided guidance at every stage of the research process.

Each analysis conducted has been presented to colleagues within either the Education, Communication and Language Sciences or the Economics departments of the university. Discussions arising from these presentations has helped to hone the analyses. Chapter 2 was presented in an early form to the North East Speech and Language Therapist (NESLT) Symposium in June 2018. Chapter 5 has been presented to the Social inEqualities in Early Development (SEED) group at the January 2020 conference. The discussion generated led to the bandwidth for inclusion robustness check presented in the Appendices.

The general format of the thesis benefitted from the feedback received in Annual Progress Reviews from Dr Carolyn Letts and Dr James Stanfield.

My colleagues and friends Scott Kirkman, Josie Tulip and Anastasia Trębacz provided helpful comments on early stages of sub-sections of Chapters 5 & 3, Chapter 2, and Chapter 1 respectively.

1 Background

1.1 Introduction

This thesis is about the importance of early speech and language development in helping to shape the subsequent life experiences of individuals born into two birth cohorts in the 20^{th} century. What does life look like for a child with early communication difficulties? How and why are their lives different from children with typical speech and language development?

Communication with one another underpins much of our society, our communities, and our economy. Most children acquire perfectly functional speech and language skills without major challenges, and it can be easy to proceed through life without conscious appreciation of the value of these skills.

However, development of these vital skills is not straightforward or guaranteed. Over 1.4 million children in the UK are estimated to have Speech, Language and Communication Needs (SLCN) (Bercow, 2018), although estimates of prevalence are highly variable both by age and sample (Tomblin et al., 1997; Law et al., 2000; Norbury et al., 2016; Law et al., 2017a). There is a widely recognised social gradient to verbal communication skills, vocabulary, and to diagnosable speech and language disorders (Law et al., 2017a).

The early socio-economic 'word gap' (Hart and Risley, 1995) may be an obstacle to social mobility established before entry into compulsory education. The Department for Education's (2017) Social Mobility Action Plan in December 2017 made reduction of the word gap, the difference in children's early vocabulary between families of high and low socio-economic status, their number one ambition. The impact of parental income on early vocabulary has been shown to explain the income gap in later academic skills in the US (Durham et al., 2007) and academic differences established in pre-school more generally have been shown to persist throughout education (Farkas and Berron, 2004).

The Bercow Report (2008) helped to raise awareness of the "profoundly damaging" effect that struggling to communicate can have on an individual. Whilst there is significant evidence of difficulties arising from SLCN in childhood, less is known about negative effects later in the life course. The few studies that have been conducted have supported the notion that early SLCN is linked to adolescent and adult outcomes including employment prospects, mental health, psycho-social outcomes and academic attainment (Law et al., 2009; Schoon et al., 2010a; Conti-Ramsden et al., 2013; Durkin et al., 2009).

Increased availability of large birth cohort datasets in many countries have allowed for a wider scope of research outcomes and more sophisticated modelling approaches within child development generally, including within speech and language sciences (for examples in recent years, see Toseeb and St-Clair, 2020; Nguyen et al., 2019; Adams et al., 2020; Zubrick et al., 2015; Camacho et al., 2019; Conti et al., 2019; Ibanez et al., 2015; Taylor et al., 2019). However, this increasing influence of longitudinal population studies has not been met with an equivalent increase in research extending beyond adolescence.

Bercow: Ten Years On (Bercow, 2018), a follow up to the original Bercow Report as the name implies, reports that awareness amongst both the public and decision makers of SLCN is still not sufficient. In the context of limited government funding, demonstrating the true cost of initial disadvantage has never been more important in supporting the argument for funding children's services and interventions.

1.1.1 Aims of the thesis

This thesis is motivated to highlight social inequalities which emerge early in life and formally test the persistence of consequences across the life course. The work presented intends to add substantial value to the field of speech and language sciences by satisfying three broad aims:

1) To utilise historical data to highlight the enduring impact of early of early speech and language development over the life course.
2) To expand the range of econometric methodologies and outcome measures used in the field to increase understanding of the breadth of these impacts.
3) To make appropriate, evidence-based recommendations for research and clinical practice.

Through these aims, this thesis increases our understanding of the long term personal and economic effects associated with early communication difficulties. Understanding of the research context is essential to this contribution and the fulfilment of all three aims. The following literature review will discuss the complexities of speech and language development including its measurement. Whilst there are many known outcomes associated with SLCN extending into early adulthood, significant unanswered questions about the life experiences of individuals with SLCN are highlighted. The review of the literature further evaluates common gaps in the sources of data and analytic methodologies. Each subsequent analysis chapter targets specific research questions arising from this review which contribute knowledge to the field. Through thorough analysis addressing these targeted research questions, the author intends to satisfy the three broad aims of the thesis.

1.1.2 Terminology

This thesis draws on multiple disciplines and decades of research. Technical terms and acronyms or abbreviations in the literature review or analyses chapters are defined at the time of their first use as well as the thesis Glossary.

1.2 Literature Review: The Development of Speech and Language

1.2.1 Communication in the Early Years

Communication is a vital part of being a person. Learning to walk and talk at a young age are the key indicators of burgeoning independence and developing the ability to communicate with others is an essential part of child development. Most children develop the ability to interact verbally with others in the first few years of life without difficulty and these skills continue to progress in complexity as the child ages (Honig, 2007). However, this ability is not equally effortless for all children. Speech, Language and Communication Needs (SLCN) is the broad term used in this thesis to refer to a wide range of potential difficulties children may be experiencing in their communication, as has commonly been used in prior UK literature (Bercow, 2008; Law et al., 2017a).

'Speech' represents the physical process through which vocal organs produce sound to convey meaning (Reilly et al., 2015) and includes the articulation of sounds, the control of the voice, and the fluency or rhythm of speech (American Speech-Language-Hearing Association, 2020). A child with a speech sound disorder may substitute or omit consonants with which they have difficulty where other children would not. Some consonant substitutions or omissions may be characterised as 'delayed', i.e., they are made by lots of children but would typically have resolved by 90% of children by a given age. Other may be characterised as 'disordered', i.e., they are not made by 90% of children at any age (Morgan et al., 2017). Other forms of well-known speech sound disorders include stammers or lisps.

David Crystal (1997) defines language as the "systematic, conventional use of sounds, signs, or written symbols in human society, used for communication and self-expression". A child with a language disorder may talk less than their peers, have difficulty understanding or remembering words (particularly those with more than one meaning) and instructions, and/or have more general difficulties with interacting socially with others such as teachers and peers (see Bishop et al., 2017 for a list of clinician-reported "red flags" for language disorders requiring intervention).

Evidence, discussed in more detail later, suggest that speech and language disorders impact a child's likely academic attainment and level of socio-emotional or behavioural difficulties (Durkin et al., 2009; Conti-Ramsden et al., 2013; Beitchman et al., 2001).

1.2.2 The History of Speech and Language Research

This thesis utilises the earliest population study to contain data on both speech and language (Newcastle Thousand Families Study, 1947: Morley et al. 1954; Morley, 1965; Miller et al., 1960), recovering data which had not been used since its initial collection. The data offers a unique opportunity to longitudinally map the outcomes for individuals with SLCN into later adulthood than has been previously attempted (as far as age 62) in the literature yet requires understanding of the context.

Speech and language were not clearly distinguished in the 1940s and this lack of distinction is reflected in the parameterisation of the data in Chapter 2. Elements of phonology and semantics, including both expressive and receptive language, are evident within the data without the modern understanding of the separation of these difficulties. The College of Speech Therapists, for which Dr Morley FRCSLT OBE was a founder and later President, was only renamed the College of Speech and Language Therapists in 1991 and the Royal College of Speech and Language Therapists in 1995. The naming of the profession mirrors the gradual separation of speech and language in our understanding of diagnosis over time. This historical context influences the necessary literature and discussion of the analysis conducted. The development of oral communication was typically viewed through an organic, biological and/or medical perspective in this era (Ingram, 1957). These frameworks remain common today, with much of speech and language pathology rooted in medical aetiological thinking (Bowen, 2009; Waring and Knight, 2013) and a tendency towards individual therapy services (Llewellyn and Hogan, 2000; Law et al., 2013). However, there have been advancements in other competing theories relating to social factors and environment, including evidence from population studies influencing diagnostic and intervention policy (Law et al., 2013) and attempts to better categorise of disorders with unknown causes (Waring and Knight, 2013; Shriberg et al., 2010; Dodd, 2005).

Therefore, this section begins with a summary of the state of the science with respect to speech and language pathology at that time, as expressed by Dr Morley, the lead speech therapist associated with the Newcastle Thousand Families Study. This is supplemented with a comparison of relevant terminology between 1954 and research as recent as 2017 to aid compatibility of the thesis with modern research.

Morley et al. (1954) summarise the five broad categories relating to disorders of “speech” in childhood. Modern categorisations of SLCN subgroups, often distinguished by the presence of a known cause, remain highly comparable with these original classifications. Much of the terminology in this thesis follows the recommendations of the CATALISE Consortium (Bishop et al., 2016; 2017)) two-phase Delphi consensus study. This means interpreting the terminology of the era using categorisations in modern diagnostic and theoretical frameworks – Table 1.1 demonstrates this.

Whilst terminology may change, the fundamental principles of human communication have not. The previous 70 years of research have increased our collective understanding of the complexities within speech and language development but the common difficulties children are reported to have experienced in their early development of communication remain recognisable.

For example, ‘dysphasia’ is commonly referenced in Morley’s early work. Whilst ‘developmental language disorder’ is the most common form of SLCN researched today, the term ‘developmental dysphasia’ was considered as an alternative label when this diagnostic term was coined through the CATALISE consortium (Bishop et al., 2017; Bishop et al. 2016). Further historical comparisons are drawn in Table 1.1 to aid comparability and highlight the lasting relevance of seminal work.

Table 1-1 Historical and modern categories of speech and language disorders

Comparison of historical terminology for main categories of speech disorders related to the data analysed in Chapter 2, continued over two pages.

Morley et al. (1954) categories of disorders of speech.	**Modern equivalent**
Disorders from deafness. *Includes both congenital nerve deafness and acquired nerve deafness.* **Dysarthria, or defective articulation.** *Due to anatomical lesions of the speech organs, to neurological maldevelopment or without abnormal neurology elsewhere.*	**Disorder associated with [primary condition X].** *Includes language impairments or speech sound disorder associated with hearing loss (sensory deficit). Speech sound disorders may be caused by 'oral structure defects' (e.g., cleft lip/palate) (Ruscello, 2008). Speech sound disorders without a known cause remain the majority, despite no clear diagnostic consensus (Waring and Knight, 2013).*
Dysphasia. *Delayed development of language. May be expressive, receptive, or global (both expressive and receptive). May be with or without a known aetiology.*	**Language Disorder** *Includes Developmental Language Disorder where aetiology is unknown. Diagnosis tends to be age > 5;0 due to potential spontaneous resolution (delay vs. disorder). Expressive, receptive, and global distinctions remain relevant. This is the most common form of SLCN.*

Morley et al. (1954) categories of disorders of speech.	**Modern equivalent**
Dyslalia. *Development of language outstrips the development of articulation resulting in the transient substitution and omission of consonants.*	**Phonological difficulties** *Dodd (2005) reports three common SLCN subgroups containing phonological deficits – delayed, atypical, and inconsistent. These remain common subsets of SLCN. 'Dyslalia' is no longer a diagnostic term in the UK but is in other countries (e.g. Hassan et al., 2020).*
Stammer. *Involuntary pausing or repeating of sounds.*	**Fluency Disorder.** *Repetitions, prolongations, and blocks are the most common disfluencies associated with stammering.*

Modern studies in the field of speech and language sciences tend to separate oral language into five main skills: phonology (the system of sounds creating a language), morphology (the structure and roots of words in a language), semantics (the application of meaning to words, including expressive and receptive vocabulary), syntax (the structure of words in a sentence, including grammatical accuracy), and pragmatics (the social use of language including norms in interaction with others, such as turn taking). Each of these skills has precursors in infancy and develop in complexity as the child ages (Honig, 2007). Whilst 'language' was not directly referred to, elements of phonology, semantics and syntax were reflected clearly in the data records in the Newcastle Thousand Families Study.

Whilst there is greater awareness and appreciation of many different childhood needs in modern schools than in previous generations, Bercow (2008; 2018) concludes that children with SLCN are still not receiving the necessary support to thrive in schools and to engage in positive life trajectories. To accurately measure the scale of this issue, two key factors (discussed in detail in the next two sections of the literature review) must be estimated:

1. The level of need in the population, i.e., the prevalence rate (number of cases in a given population) across various ages in childhood in the UK.
2. The economic cost associated with SLCN on an individual and societal level.

1.2.3 The number of children with SLCN

1.2.3.1 Inclusivity

Whilst some research papers exclude children with primary conditions, this thesis includes children with recognised causes of SLCN for three reasons:

1. These children are a relevant portion of those disadvantaged by service provision (Bercow, 2018) and show no difference in socially valued outcome to DLD/SLI individuals (Tomblin et al., 2008). The categories described in Table 1.1 primarily focus on medical aetiologies, suggesting these will make up a high proportion of observed SLCN cases in Chapter 2.
2. Recent key papers recognise the value of inclusivity in research and clinical access to intervention (Bishop et al., 2017; Norbury et al., 2016).

3. Data available containing life course outcomes is unlikely to have adequate provision of data on all potential causes of SLCN due to its age. Other measures of cognition are included as control variables where available.

This inclusive principle aligns with the core aims of the thesis outlined in Section 1.1.1 to understand life experiences. Furthermore, the use of control variables rather than exclusionary criteria helps to maximise the explanatory value of the data available by isolating the impacts of specific skills.

Speech, Language and Communication Needs will be therefore used where the data is available in Chapter 2. This term is most used in relation to special educational needs policy in the UK, and therefore is less concerned with aetiology than the previous diagnostic frameworks references for disorders. As special needs were not conceptualised in the same manner prior to the Warnock Report 1978, the measure used includes of less severe cases of SLCN than would have been diagnosed with a formal disability in the 1950s.

From Chapters 3 to 5, an assessment score measuring receptive vocabulary is used as the indicator of language ability. Whilst language disorders are a subset of SLCN (see Table 1.1) often diagnosed using similar assessments, the analysis within this thesis does not intend to identify or label groups as having or not having language disorders. Rather, the measure is used to indicate a spectrum of ability at age 5 and to highlight the differences that emerge between children across the distribution of early childhood vocabulary.

1.2.3.2 Prevalence estimates

Tomblin et al. (1997) conducted a wide-scale and long-running analysis of Specific Language Impairment (SLI) prevalence from a population sample of 5- to 6-year-olds in the US. SLI here means a language disorder without a primary aetiology and is comparable to the modern diagnostic term of Developmental Language Disorder (DLD, as mentioned in Table 1.1) within this age group. They estimate that 7.4% (95% Confidence Interval (CI): 6.4-8.5%) of children in this age range and sample were diagnosable with SLI and found this to be consistent with the literature of smaller studies conducted prior to this key paper.

7218 children were screened and 26% (n=2084) were followed up for a range of measures in order to identify SLI. The criteria for diagnosing SLI was dependent on two criteria:

1. Scoring lower than 1.25 standard deviations below the mean on at least two of seven different language tests.
2. Scoring higher than 1 standard deviation below the mean on a performance IQ test.

Together, these two criteria were used to identify the language impairment, and the lack of a primary condition. The present thesis, as stated, does not depend upon the lack of a primary condition although the work of Tomblin et al. (1997) is still relevant for estimating the need of one subset of all Individuals with SLCN at a given age and for describing the typical methodologies for identification which have been influenced by this work.

Law et al. (2000) utilised a meta-regression approach to summarise the prevalence literature of those with primary speech or language delays (i.e., no primary condition or aetiology, but not restricted to language). They find that prevalence estimates of speech and/or language delays range from 2.0% to 24.6% dependent on the nature of the difficulty and the age of the prevalence estimate. Law et al. (2000) report the median estimate of primary language delay at age 5;0 to be approximately 6.8%, and the estimated median prevalence of primary speech and/or language delay at 5;0 (five years and zero months of age) to be approximately 11.8%. This shows that this methodology broadly supports the findings of Tomblin et al. (1997) and that the findings of Tomblin et al. (1997) would underestimate the required provision of services through exclusion of those with speech difficulties.

The most recent prevalence estimate in the UK literature was conducted by Norbury et al. (2016) from a population sample of 7267 four to five-year olds in Surrey. Norbury et al. (2016) describe their identified group as having a “language impairment severe enough to hinder academic progress” and reports findings for those both with and without a primary condition. They find a prevalence rate of 7.6% (95% CI: 5.3-10.7%) for children without a primary condition which is remarkably comparable to the findings of Tomblin et al. (1997). They find a prevalence rate of 9.9% (95% CI: 7.4-13.2%) for children either with or without a primary condition. They state that this finding identifies the additional need for service provision from children with primary conditions who would benefit from language interventions. Only 3.9% of the children identified with DLD in Norbury et al. (2016) had a statement of special educational needs, and only 39% of the children identified with DLD received a language intervention outside of school.

Estimated derived from community cohort studies suggest a prevalence around 3-5% of four-year-olds with speech sound disorders with no known cause (Wren et al., 2016; Eadie et al., 2015; Reilly et al., 2015). The methodologies used to estimate prevalence of speech disorders are typically similar to language disorders, albeit with different standardised assessments. Cut-off scores and battery testing are combined in Wren et al. (2016), with this identification method used in multiple papers.

Law et al. (2004)'s review of prevalence estimates demonstrates the significant overlap between children with speech and with language disorders. Therefore, overall SLCN prevalence cannot be determined from a simple addition of these estimates. This is unsurprising given the feedback loop between speech and language in early child development. Many children referred to speech and language therapists would be identified with SLCN including elements of both speech and language difficulty.

In 2004, approximately 85-90,000 children between the ages of two and six were referred to Speech and Language Therapists in the UK (Broomfield and Dodd, 2004). The closest time period with available UK population statistics (Office for National Statistics, 2019) show that in 2008, approximately 3,518,930 children lived in the UK between the ages of two and six. This suggests a prevalence rate, as measured by referrals, of around 2.5% which is below even the lowest prevalence estimates available within the literature, including the estimates available in 2004. Lindsay and Strand (2016) report that approximately 1.6% of children in schools (age five to 16) were identified with SLCN as their primary special educational need in 2011, although this represented a significant rise from 0.9% in 2005. Underdiagnosis has been noted in the literature relating to retrospective oral language assessments of children in youth offending institutions (e.g. Gregory and Bryan, 2011). Underdiagnosis appears common given the difference between research-estimated prevalence and clinical referral rates (McGregor, 2020). The constraints of resources in speech and language therapy is likely to be one factor, with a third of children waiting over a year to access Speech and Language Therapy (Bercow, 2018). The limitations of the referral process, reliant on teacher or parent observations in the absence of an accurate diagnostic screening process, are likely to be another (McGregor, 2020; Bishop et al., 2017).

The above papers demonstrate a degree of consensus, that SLCN in its various forms will impact upon a sizeable minority of children. Many population estimates occur at approximately age 5 due to the relevance of school entry in both identification and impact of

SLCN. Approximately 10% of children (Law et al., 2017a) or over 1.4 million children (Bercow, 2018) experience SLCN. The Norbury et al. (2016) statement that “approximately two children in a classroom of 30” experience language impairment severe enough to justify additional support has gained some media attention (The Guardian, 2017) and whilst there are issues with prevalence estimate methodologies (discussed next and analysed in chapter 5), the scope of the national problem of SLCN is clearly worthy of considerable attention.

1.2.3.3 Threshold values on standardised assessments

Standardised assessments form a key part of the methodology for estimating prevalence in a given population as well as for clinical diagnosis (Spaulding et al. 2006; Plante and Vance, 1994; Law et al., 2017a, Le et al., 2020b). These assessments are intended to identify a disorder as opposed to simply provide a ranking of abilities, as is more common in educational assessments (Spaulding et al., 2006). Threshold values are applied to the distribution of scores on a test in order to maximise ‘sensitivity’ and ‘specificity’, the diagnostic accuracy of the test in identifying those with the disorder as being those who fall below the test threshold.

Sensitivity represents the percentage of individuals who truly have the disorder who are correctly identified by the test’s threshold. High sensitivity suggests a high rate of true negatives.

$$\text{Sensitivity} = \frac{\text{N (With Disorder and a score below the threshold)}}{\text{N (With Disorder)}}$$

Specificity represents the percentage of individuals who do not have the disorder who are correctly identified by the test’s threshold. High specificity suggests a high rate of true positives.

$$\text{Specificity} = \frac{\text{N (Without Disorder and a score above the threshold)}}{\text{N (Without Disorder)}}$$

A threshold is chosen to achieve high rates of sensitivity and specificity, usually requiring a minimum of 80% sensitivity and 80% specificity, as first recommended by Plante and Vance (1994). The threshold chosen is typically at 1, 1.5 or 2 standard deviations below the mean, noted by Wiig et al. (1992) as representing mild, moderate, or severe levels of disorder on language assessments. A threshold at one standard deviation below the mean, will typically have a high sensitivity and low specificity through estimating a higher percentage of the distribution having the disorder, which may be preferable for screening procedures. A more severe threshold, by comparison, increases specificity and decreases sensitivity. This is more suitable for situations where the cost of falsely identifying a disorder increases, relative to the cost of falsely failing to identify a disorder.

Law et al. (2000; 2017) note the widespread presence of common threshold values applied to language tests (also at 1, 1.5 or 2 standard deviations below the mean) for estimating prevalence. The choice of threshold clearly impacts the resulting prevalence estimate. If test scores follow a perfectly normal distribution then a mild threshold implies a 15.9% prevalence, a moderate threshold implies a 6.7% prevalence and a severe threshold implies a 2.3% prevalence.

The distribution of scores on standardised assessments within children often involve some clustering at the bottom and top ends of the distribution, also known as floor and ceiling effects, whereby scoring 0% or 100% is relatively more likely than scoring 1% or 99%. This explains the prevalence rates within Law et al. (2017a) tending to be slightly above those implied by the normal distribution, as children may cluster below the cut-off point.

Inclusion of more than one test score criteria, as in the seminal work by Tomblin et al. (1997), is observed to reduce the prevalence estimate relative to a study utilising an equivalent score on one test (Law et al., 2000) which is a natural effect of the child having to “fail” two tests rather than just one to be identified in the given process.

Awareness of these issues around prevalence estimates being dependent on the cut-off scores has clearly been noted for at least 20 years and remains a topic of discussion in recent research (Le et al., 2020b). In 2017, a report by Law et al. (2017a, p17) for the Early Education Foundation summarises prevalence estimates with clear description of the thresholds applied by each paper reported.

Whilst thresholds may be useful in optimising a trade-off between sensitivity and sensitivity, they are not designed to provide accurate prevalence estimates. Despite this, a significant number of prevalence estimates in the literature utilise this approach (see Law et al., 2017a, p.17 for a summary relating to language assessments).

1.2.4 Stability of Speech and Language Development Profiles

The prevalence of SLCN within different age ranges varies dramatically across childhood as age norms change. Law et al. (2017a, p17) report the median prevalence estimates across various ages in childhood. These are typically highest in younger age groups as many children spontaneously resolve their early difficulties whereas fewer difficulties emerge after school entry. Reilly et al. (2010) for example show that 70% of 'late talkers' at age two are resolved by age four, whereas McKean et al. (2017a) demonstrate within the same cohort that the vast majority of children with language difficulties at age four are unlikely to resolve by age 11. Morgan et al. (2017) found that four-year-olds are significantly more likely to resolve their speech errors if they are classified as delayed (e.g., a consonant substitution that is common in younger children, but 90% of your age peers no longer make) with 70% resolving than if they can be classified as disordered (e.g., a consonant deletion that 90% of children of any age wouldn't make), for which only 40% resolved.

As prevalence refers to the number of cases within a given population, this tendency for resolution or persistence within different age groups suggests that the total number of children who experience SLCN at some time prior to adulthood is higher than the prevalence estimate within the age range with highest prevalence, typically around three years of age (Law et al., 2017a).

The potential for spontaneous resolution of both speech and language disorders in early childhood has been recognised for many years (see Morley, 1965, although the types are not distinguished into speech and language) and holds within longitudinal, population samples (Ghassabian et al., 2014; Reilly et al., 2010; Wren et al., 2016). This causes long-standing complications for policymakers determining the optimal timing for effective intervention. Screening for language difficulties before school entry can be costly and results tend to be weakly predictive of later abilities (Law et al., 2017c; Reilly et al., 2010; Dockrell and Marshall, 2015).

Most resources are targeted at primary school age children (Lindsay et al., 2012), and prevalence of identified SLCN does indeed fall from age five to 15 (Lindsay and Strand, 2016). However, most children with deficits in language or speech at school entry will continue to have deficits relative to their peers (McKean et al., 2017a; Bornstein et al., 2016; Bornstein et al., 2018; Wren et al., 2016). This suggests that the identification of SLCN in schools does not necessarily reflect age norms, and that interventions may be targeted at children after their trajectories of language development are comparatively stabilised.

A small group has been identified in previous research whose language difficulties emerge after school entry (Law et al., 2012a; Law et al., 2017a) with difficulties demonstrated in school-age activities such as reading (Snowling et al., 2016). This group are relatively poorly researched as they would not be contained in clinical samples recruited prior to the onset of their difficulties, and they represent only a small number in longitudinal population studies.

The effects of SLCN into adulthood have been shown to differ by language profile, in that more severe negative outcomes tend to be observed for individuals whose SLCN persists rather than resolves. Stothard et al. (1998) found that language difficulties resolved by age 5;6 score equally well to control groups by age 16 on measures of oral language and literacy skills. Those who did not resolve by this time, however, fell further behind on vocabulary tests over time. Snowling et al. (2001) demonstrate that persistence of language difficulties within school years predicts performance in academic examinations at age 16 relative to resolving language difficulties.

One source of instability in profiles of speech and language in research is the use of strict identification thresholds on standardised assessment tests. Children who fall very slightly below a threshold may be shown to 'resolve' if they score only slightly higher along the distribution at a later timepoint and vice versa. Given the presence of measurement error on any assessment, this may be picked up by a predictive model as resolving or emerging difficulties where no real change has occurred in the child. It is important to separate this technical instability at the margin from genuine changes explaining differences in later outcomes.

In addition to the issue of general measurement error, a process known as 'regression to the mean' causes similar issues of apparent instability, which is not reflective of a child's underlying, unobserved communication ability. If a child scores within the lower end of the distribution at the initial time point, it is more likely that they have been mis-measured negatively than positively. Therefore, in repeated measures with an expected error term of zero, their position on the distribution is likely to appear to improve. This has been demonstrated to confound analyses of social mobility within children's cognitive scores (see Feinstein, 2003; Jerrim and Vignoles, 2013). Feinstein (2003) showed a crossover effect, whereby initially high ability, low SES children were outperformed by initially low ability, high SES children. Using simulated data with randomly drawn measurement error and time-invariant mean values, Jerrim and Vignoles (2013) were able to demonstrate that the same crossover could be observed purely through regression to the mean. Whilst this thesis does not model trajectories of ability using distributional tests, this issue is relevant to the estimation of SLCN prevalence at various ages and the scale of the problem which this thesis is motivated to address.

Regular assessment reduces the risk of mismeasurement. Where multiple measures of language are taken longitudinally, statistical techniques which are less susceptible to small movements across thresholds may be used. Latent Class Analysis, for example, has been used to characterise subgroups of children by their trajectories of vocabulary development (McKean et al., 2017a). Profiles of decline, improvement or stability are observable. Where longitudinal data is unavailable, assessments made before school entry are poorly predictive of later ability.

Having multiple measures taken cross-sectionally may also reduce the risk of mismeasurement from a single test. Latent techniques have identified a core language ability drawn from multiple facets of language, shown to be stable into adolescence (Bornstein et al., 2018). However, there is a risk that the composite scores of multiple measures are often not tested for their reliability in making accurate diagnoses (Nitido and Plante, 2020).

The importance of gathering multiple assessment scores, either over time or cross-sectionally, has gained recognition as the field has developed over time. Therefore, older cohorts (which are necessary to capture life course outcomes) do not always contain either longitudinal or varied cross-sectional assessments. Where multiple assessments are not available, measures of receptive language are more representative of a stable core language ability in terms of their stability over time and cross-sectional correlation with other facets of language (Dionne et al., 2003; Dale et al., 2010; Bornstein et al., 2018; Nickisch and von Kries, 2009; Law et al., 2004).

These examples demonstrate the value of a broad range of analytic techniques, alongside longitudinal study designs, for enhancing the evidence base and understanding of speech and language development.

1.2.5 Associated Risk Factors

There are multiple early life risk factors shown to predict higher levels of SLCN within specific populations. Multiple factors may interact in a child's life, and it is rare for any of these risk factors to be individually responsible for a childhood condition (Law et al., 2013). However, each of the following risk factors are relevant to the current thesis and discussed as follows:

- Socio-economic status
- Gender assigned at birth
- Cognitive ability
- Birth weight
- Birth order and number of siblings

Each of these risk factors are summarised in relation to SLCN below. SES and gender are crucial to understand in this thesis, as they are consistently huge predictors of many relevant life outcomes across their changing historical context over the 20th century. Implications of this research would be interpreted differently for different groups and understanding these differences is vital for maximising the benefits for research and clinical practice.

1.2.5.1 Socio-economic status

There is a well-established social gradient to disorders of communication. For example, both SLCN, and DLD specifically, have been recently posed as a public health problem due to their unequal burden across society (Beard et al., 2018; Law et al., 2017b). This social gradient is reflected in the prevalence estimates observed in different areas using the same norm-referenced standardised assessments. For example, the recent UK estimate by Norbury et al. (2016) discussed in the previous section is recognised by the authors as potentially underestimating the wider UK population level prevalence due to it being obtained from a sample in Surrey, a relatively advantaged area of the UK.

Locke et al. (2002) find that over 50% of a sample drawn from a socially disadvantaged areas, measured by the proportion of free school meals in the local area, scored below the nationally referenced 16th centile, approximately 1 standard deviation below the mean, on overall language scores (expressive and receptive) on the widely used Clinical Evaluation of Language Fundamentals: Pre-school test, or CELF-P UK (Wiig et al. (1992). This is an extremely high prevalence rate relative to those described in the previous chapter, although it is notable that this is a relatively mild threshold and that the free school meals proportion was not reported in order to identify the severity of social disadvantage. Law et al. (2011) find relatively similar results to Locke et al. (2002), that 39.9% of children experience diagnosable language impairment within a sample drawn from the most deprived quintile of Scotland using the Scottish Index of Multiple Deprivation. Letts et al. (2013) show that whilst there is a link between neighbourhood-level social disadvantage and language skills, they did not find as severe a correlation as that observed by Locke et al. (2002) and Law et al. (2011) despite using an equivalent Index of Multiple Deprivation measure to Law et al. (2011) for the nursery the children attended.

Whilst this social gradient is also demonstrated across income quintiles within the Millennium Cohort Study, although there is a high level of variability within each quintile (Law et al., 2013; Wareham et al., in preparation). Each income quintile includes children from both the top and bottom of the language distribution and the social gradient is best observed using central tendencies (mean, median and mode) within each quintile.

The underlying mechanism behind the correlation of language and socio-economic status is contested. Input from parents in early years has been consistently shown to influence human capital generation more generally, including measures of communication ability (Cunha et al., 2010; Almond et al., 2018). Whilst social disadvantage is commonly observed in cognitive skills generally from a very young age (Feinstein, 2003; Jerrim and Vignoles, 2013), Locke et al. (2002) demonstrate that language scores are significantly lower than general cognitive measures in their sample of children from areas with a high proportion of free school meals. The finding that the social gradient associated with language is more severe than with other cognitive skills is supported by the Department for Education (2017).

A landmark study by Hart and Risley (1995) details one potential mechanism relating to parental input known as the 'big word gap' or the '3-million-word gap'. Analysing a sample of 42 children from three parental occupational statuses (white collar, blue collar, welfare) over the course of the first three years of life, they compare the quantity of input the children receive in one-hour samples of parent-child interaction. They conclude that children from the most advantaged families, as defined by parental occupational status, were exposed to a much larger range of vocabulary than children from the less advantaged families. This measure of advantage was strongly related to the education and income levels of the parents.

The specific size of the word gap was reliant on a high level of extrapolation within a small sample and none of these measures of socio-economic status would directly cause differences in maternal response to the child's speech. However, Hoff (2003) demonstrated the exact same effect when comparing 33 high-SES to 30 mid-SES families, as measured by both parental education and employment, and found that all SES differences in expressive vocabulary growth were explained by the observed differences in maternal speech. Furthermore, Hirsh-Pasek et al. (2015) have shown that both quantity and quality (for example, child-directed speech as opposed to overhearing) of the child's exposure to words help to explain these associations. The home environment generally explains many of the socio-economic associations, particularly for receptive difficulties, although expressive difficulties do have a stronger correlation to socio-economic status (Lohndorf et al., 2018).

Sullivan et al. (2017) document similar findings relating to parental education and occupational status within the Millennium Cohort Study (CLS, 2000), a UK population-level sample. They find that the main mediator of parental education predicting child vocabulary is parental vocabulary levels and that this is particularly important for predicting language change between five and 14 years of age. This is supportive of the original finding in Hart and Risley (1995), that one key explanation of the large social gradient in child language relative to other cognitive skills is likely to be the exposure to varied language that the child is subjected to across childhood. Letts et al. (2013) found that maternal education was a significant predictor of children's expressive and receptive vocabularies, referred to as production and comprehension, within a large standardisation sample. This effect was strongest for younger children with mothers who had the minimum number of years of education, i.e., they left school as soon as it was no longer compulsory to attend. Children within this group were over 1.6 times as likely to fall below the normal range (below the 16^{th} centile) relative to the whole sample.

Differences in the quality of input received by the child may not be causally related to the parents' vocabulary. A higher parental vocabulary level suggests economic advantage, which suggests lower stress levels within the family, increasing the available time and financial resources available for positive child-centred behaviours or activities (Conger et al., 1994; Heberle and Carter, 2015). A different environment for the parents, in terms of economic resources and stress, could therefore theoretically translate into improved home learning environments for the children. For example, Wren (2016) demonstrate that home ownership was a significant factor in predicting the prevalence of persistent speech sound disorders, which constitutes a substantial proportion of clinical caseloads. The parental environment factors (such as income, education, access to resources locally, stress and nutrition) are often regarded as distal, whilst language exposure is a proximal predictor, shown to impact brain function directly (Romeo et al., 2018). Economic policy is required to target many distal factors, whereas proximal factors are often more directly targetable for local practitioners and teachers.

The National Literacy Trust (2018) published a report on improving the home learning environment, suggesting “play, chat, read” as a slogan for families. This report acknowledged the presence of significant barriers to language learning within these homes, including factors such as time, financial resource, and stress. Furthermore, there was an acknowledgement of the reduced availability of neighbourhood resources (libraries, parks, safe areas) within more deprived areas. Reducing the socially graded word gap observed in studies such as Locke et al. (2002) requires policies addressing area-level disadvantages as well as factors within the home.

Reilly et al. (2010) find that a variety of maternal factors correlated with socio-economic status predict the language ability of children at four years of age including maternal mental health, vocabulary, education, and age at the child’s birth. These are factors often measured in social surveys which may proxy for the underlying mechanisms such as the range of vocabulary the child is exposed to. There is significant unexplained variance in children’s language at age four within these predictive models, potentially reflecting the relatively poor potential for any measure of socio-economic status to capture a child’s experience at an individual level. National longitudinal studies are unlikely to measure direct maternal responsiveness on an individual level and note the specific behaviours, such as using expansions of the child’s verbalisation to encourage improving vocabulary in slow-to-talk toddlers (Levickis et al., 2018a; Conway et al., 2018).

Letts et al. (2013) report that the impact of maternal education is strongest in younger children for expressive and receptive vocabulary. Most of the differences they observe are reduced through the first years of schooling. This finding contradicts those of Taylor et al. (2013), who found that socio-economic disadvantage does not predict vocabulary at age four but does predict vocabulary growth between age four and eight. Taylor et al.’s findings are more in line with the expectation from the child health literature. Case et al. (2002)

demonstrate that as children age, the relationship strengthens between the child's health outcomes and their socio-economic status, as in Taylor et al. (2013). McLeod and Harrison find that 10 of 14 studies assessed find a significant association between parental education and language, even when language is not measured as a trajectory.

Social disadvantage may occur at a neighbourhood level, as shown in the Locke et al. (2002) study, or at a family level through parental differences, as shown in Hart and Risley (1995) and Sullivan et al. (2017), or at an individual level through factors such as general cognitive ability, as shown in Feinstein (2003). Each of these levels of disadvantage are interrelated and risk factors do not tend to occur in isolation (Taylor et al., 2013), as is recognised in some way in each of these papers.

In the absence of data on parental input or parental vocabulary levels, measures of socio-economic status (such as parental occupational status, education or income) are useful proxies for behaviours and neighbourhood disadvantages which may be unobserved. The use of proxies for socio-economic status in social science research is both necessary and limited (Goldthorpe, 2000). The crude nature of these proxies is likely to result in a relatively low proportion of variance explained in the SLCN measure, although this is often the case anyway, particularly in children as young as two (Reilly et al., 2010).

Blanden (2006) analysed a subset of children disadvantaged by poverty in childhood within the British Cohort Study 1970. A range of childhood performance measures at age five and ten were tested as predictors of upwards social mobility. Vocabulary at age five was found to be the strongest predictor, demonstrating the important role that language plays in the intergenerational transmission of advantage.

To accurately measure the impact of SLCN over the life course, inclusion of measures of socio-economic status is essential as socio-economic disadvantage is related to both the predictor (SLCN) and the majority of life outcomes (Marmot, 2010), with potential interaction effects between the three (e.g., Blanden, 2006).

Thus, social inequalities associated with speech and language are a statistically relevant risk factor for the analysis. More importantly, highlighting these social inequalities in early life and demonstrating their impact on further inequalities experienced across the life course is the core motivation for the whole thesis. Increasing awareness of the value in reducing early disadvantage is a theme in much of the literature reviewed and the analysis conducted.

1.2.5.2 Gender assigned at birth

Boys, as measured by gender assigned at birth, are more likely to be identified with SLCN in childhood than girls. This association is strongest within clinical samples, with Leonard (2014) finding boys twice as likely to be identified with language impairment as girls in a review of the prevalence literature. Tomblin et al. (1997), using a population-level longitudinal design, found more comparable proportions of specific language impairment in boys and girls (8% and 6% respectively).

There is evidence that the different referral rates by gender may be related to the higher likelihood of "acting out" and exhibiting externalised behaviour problems amongst boys (Beitchman and Brownlie, 2013). Law et al. (2009) demonstrate that whilst girls performed worse on average on the English Picture Vocabulary Test within the British Cohort Study (1970) at age five, boys were in fact still more likely to be currently receiving intervention. This could suggest a selection bias in referral rates for equivalent levels of difficulty, as hypothesised by Beitchman and Brownlie (2013). Despite this potential treatment bias, boys are more likely to experience persistent speech sound disorders (Wren, 2016).

This could alternatively reflect a gender imbalance in forms of SLCN unrelated to the English Picture Vocabulary Test, a measure of receptive vocabulary. Harrison and McLeod (2010) report that males are almost twice as likely (1.97 odds ratio) as females to present with expressive language difficulties. The odds ratio for receptive difficulties, whilst still more common in males, is lower (1.39 odds ratio). Peckham (1973) report that stammers were twice as common for boys than for girls within the National Child Development Study, a nationally representative longitudinal study design. These findings would not totally account for the different ratios of specific language impairment found between Leonard (2014) and Law et al. (2009) but may explain some of the noted referral bias.

The most recent estimates in Norbury et al. (2016) find a gender ratio of 1.22:1, with boys slightly more likely to present with language disorders, although this difference is statistically non-significant. However, the gender ratio for language disorders with primary conditions is 3.3:1 and the ratio for the limited screening procedure is 2:1. This may help to explain why the ratios observed within clinical settings are far stronger than those observed in isolated test conditions for population-level samples.

Within the Department for Education's report on the Early Years Foundation Stage Profile statistics release, 66.4% of boys relative to 79.1% of girls achieved the expected level of Communication & Language and Literacy performance (National Statistics, 2018). This gap of 12.7 percentage points is decreased from 16.7 percentage points in 2013, the first year of the report, although remains a substantial difference worth controlling for in analyses.

As with socio-economic status, gender is strongly related to three core constructs within this thesis: economic outcomes, the generational context, and the implications for research and clinical practice.

1.2.5.3 Cognition

Speech and language development are correlated with many other indicators of child development. Communication more generally is commonly included in many cognitive and non-cognitive composite scores (Feinstein, 2003; Cunha et al., 2010; Conti et al., 2019; Almlund et al., 2018; Heckman and Kautz, 2012).

Since Spearman's 1904 paper on generalised intelligence 'g', often parameterised as IQ in more recent decades, researchers have shown awareness of the clustering of cognitive abilities in individuals (for a more recent example, see Blanden and Machin, 2010). Early measures of cognitive ability often rely upon Executive Function tasks or upon reasoning, inference and pragmatic skills which are all related to early language abilities in children (Law et al., 2017a).

Processes of human capital accumulation are endogenous. Genetics and environment impact on early neurobiology and cognition. Cognition and language ability interact in their ongoing development. For example, working memory is shown to enable receptivity to parental input for initial language acquisition (e.g. Fyfe et al., 2019), whilst language enables reading and writing (e.g. Snow, 2020). Exposure to books then increases children's vocabulary and their access to school curriculums and so on. This explains the higher return to parental and policy investments early in childhood relative to later (Cunha and Heckman, 2006).

Most commonly, research focuses on the link between language and literacy (Snow, 2020; Im-Bolterm and Cohen, 2007; Law et al., 2009) given the implicit relationship between reading, writing and oral language. However, associations between language and cognitive ability are demonstrated for less obvious relationships. Children with speech difficulties at age four to five have been demonstrated to score below typically developing matched control groups on measures of reading, spelling and maths at age seven (Nathan et al., 2004). This finding was particularly pronounced for those whose speech difficulties had persisted between age four and seven. This demonstrates that the correlational effects of SLCN are not limited to the school subjects with obvious overlap. This correlation of communication with mathematical ability has extended to children with language disorders, not just speech, with working memory acting a partial mediator within this population (Fyfe et al., 2019).

Many research papers focus on language disorders in the absence of cognitive deficits or primary aetiologies, usually using a threshold value indicating normal range IQ scores relative to a threshold value indicating low language, known as Developmental Language Disorder (DLD) or previously Specific Language Impairment (SLI). This thesis does not exclude participants on the basis of IQ below the normal range. Instead, this thesis controls for the available measures of cognitive ability to increase the accuracy of the estimated effects of SLCN in childhood. Accounting for these factors is important because cognitive ability and educational attainment have been shown to predict adult socio-economic outcomes (Heckman and Mosso, 2014).

1.2.5.4 Birth weight

Low birth weight has been identified as a risk factor for both poor speech and poor language development in childhood in a systematic review (Nelson et al., 2006). Reilly et al. (2010) show that low birth weight predicts low expressive and receptive vocabularies at age four within a longitudinal sample. Variability in the definition and measurement of birth weight makes it difficult to quantify the strength of this effect (Barre et al., 2011).

Birth weight has similarly been shown to predict cognitive ability at age nine within the Newcastle Thousand Families Study (Pearce et al., 2005) and an umbrella review of systematic reviews and meta-analyses shows evidence that early birth weight can predict some later life health outcomes (Belbasis et al., 2016).

1.2.5.5 Birth Order and Number of Siblings

Birth order is a recognised risk factor for poor speech and language development, with firstborns commonly displaying the strongest communication abilities (Reilly et al., 2007). This is usually explained using the same logic as those described previously in Hart and Risley (1995) and Hoff (2003), that higher levels of parental input result in greater communication development of the child. The presence of family resources, including time, may increase the parental availability for these language learning opportunities (Heberle and Carter, 2015) and these resources may decrease per child as the number of children in the family increases. Zubrick et al. (2007) find that having two or more siblings is a risk factor for language delay.

The quality of input the child receives may alter if they learn speech from older siblings, although Hoff-Ginsberg (1998) suggest that younger siblings demonstrate advanced conversational skills that may compensate their reduced lexical and grammatical skills. There may therefore be a differential impact of family size depending on the age of the child which could be modelled through the methods followed in Taylor et al. (2013), where risk factors are tested for their impact on vocabulary growth across childhood. As the average UK family size decreases over time (Lutz et al., 2006), the national importance of this risk factor may similarly decrease.

1.2.5.6 Risk factors unobserved in the present thesis

Family history of SLCN is a known predictor of SLCN for the child (Choudhury and Benasich, 2003; Bishop et al., 2017) with a genetic component indicated through twin studies of motor difficulties and language impairment (Bishop, 2002). This may work independently or in conjunction with the parental input and vocabulary findings outlined in Section 1.2.5.1 relating to socio-economic status. This heritability suggests a genetic component to the parental transmission of vocabulary noted above. The genetic or biological inheritance of the child interact with their early environment (Erbeli et al., 2018). This suggests malleability of outcomes for any given child through environmental changes. Given the social gradient described in speech and language development, it is relevant that environmental factors have been shown to be more predictive of outcomes than genetics in socio-economically deprived environments (Tucker-Drob and Bates, 2016). A child's home learning environment is consistently shown to predict their development of speech and language in the reviewed literature (Bishop et al., 2016; Levickis et al., 2018a; Hart and Risley, 1995; Hoff, 2003). Whilst future research and data may include more measures of genome sequencing, the data from the mid-20th century which this thesis aims to utilise for the availability of measures into adulthood does not contain this information. This does not limit the potential of this thesis to achieve its three core aims.

Similarly, bilingualism is a broad field of study within speech and language sciences. Measurement of language ability is intended to occur in a child's primary language (Bishop et al., 2017) for diagnostic and research purposes. Cultural differences can limit the potential for specific vocabulary tests to accurately measure equally across all populations (see for example the Peabody Picture Vocabulary Test: Washington and Craig, 1999). Within this thesis, the English Picture Vocabulary Test (Brimer and Dunn, 1962) excludes English as an Additional Language children and the Newcastle Thousand Families Study did not specify any cases where English was not the primary language of the child. This simplifies the issue of bilingualism which is likely to be more relevant in studies of later cohorts given the diversification of the population. This limitation in the modern applicability of the current research is considered in the discussion.

1.3 Literature Review: The Value of Speech and Language Throughout Life

The discussed literature details the complexity of early speech and language development. A substantial minority of children experience speech, language and communication needs each year with particularly high levels of reported difficulties amongst children from disadvantaged backgrounds. Whilst development of these skills is non-linear, difficulties which persist into the school years are less likely to resolve spontaneously.

Most resources for intervention are targeted within the school years, with intervention intended to ameliorate the associated negative impacts of experiencing SLCN. There are few studies examining the role of communication skills in predicting later achievement, with the exception of research focused on clinical populations, e.g., autistic children (Gutman and Schoon, 2013).

Great Ormond Street Hospital (2018) define the clinical outcomes for speech and language therapy as measurable changes in health, function or quality of life resulting from care. There is no consensus on a consistent, psychometrically sound and clinically meaningful measure of intervention benefits in the field of speech and language therapy (Mullen, 2004; Fratalli, 1998).

When parents at Great Ormond Street Hospital Speech and Language Therapy Clinic, in the two years to August 2018, asked questions of the therapists, they most commonly asked about the long term expectations for their child more often than any other topic, including the nature and cause of the difficulty, their child’s progress or more technical questions about the treatment. Functional ability of the child across the life course was the highest concern for these parents and this should be reflected in the breadth of outcomes considered in research.

Dockrell et al. (2014) summarise the parental valuations of their children's outcomes:

- Communication
- Independence
    - Academic potential
    - Personal achievements
    - Safety
    - Earnings potential
- Sociability
    - Making friends and avoiding social exclusion
    - Confidence as a consumer

The following literature review begins by discussing the approach to intervention studies in speech and language sciences relative to other areas of child development. In particular, the valuation of the benefits of interventions are discussed and the importance of outcomes into adulthood for measuring this.

The broad literature on academic achievement for children with SLCN are discussed, with focus applied to the lack of research into further education and the changing context over time of education in the UK. This leads onto the poor behavioural outcomes identified within children with SLCN, particularly within schools, and the ongoing relationship between non-cognitive personality traits and communication.

Finally, outcomes into adulthood are discussed, focusing on the key papers available and the areas where further research is necessary. In particular, the lack of research into the health outcomes and earnings potential for children with SLCN is highlighted.

Frank Field (2010) wrote extensively about the value of child wellbeing indicators in predicting later life outcomes. He utilised Max Weber's 'life chances' framework, representing the opportunities available to an individual to improve their quality of life over the life course; a phrase subsequently used throughout the time David Cameron was Prime Minister. The phrase 'chances' is neatly applicable to the statistical modelling presented in this thesis, whereby predicted outcomes do not reflect actual outcomes for all individuals, but highlight the inequalities generated by factors outside of the child's control. Johnson and Kossykh (2008) and Field (2010) agreed that child language was a crucial component in the indication of early life chances. This thesis contributes further evidence of the direct relevance of early communication in creating and grasping opportunities across the life course within wide-ranging social contexts.

1.3.1 Interventions for Speech, Language and Communication Needs

A meta-analysis by Law et al. (2004) found that interventions can be hugely successful in improving outcomes for individuals. Effect sizes were found as large as two standard deviations in language scores for children exposed to a successful intervention. However, data was lacking for children with receptive-only difficulties, partially due to a referral bias that means receptive difficulties are not proportionally represented in clinically drawn samples. It is also demonstrated within this meta-analysis that interventions which last over eight weeks are more effective than shorter interventions. This is supported by more recent evidence that intervention can be effective in improving outcomes and cost-effective (Lindsay et al., 2012; Law et al., 2012b).

1.3.1.1 Speech and Language Therapy

Speech and Language difficulties are typically tackled through psychosocial interventions with social, psychological, cognitive or educational components, intended to improve a skill or behaviour (Mol and Bus, 2011). Accessing these interventions can be difficult, with around a third of children waiting over a year to get the speech and language therapy they are referred for (Bercow, 2018).

Successfully navigating the access to services is strongly related to social advantage, with evidence of education and experience removing even the non-financial barriers to service access (Willingham, 2012). For example, when free early childhood education was made available to disadvantaged children at age two in the UK, only two thirds of places were taken up. Law et al. (2017b) suggest this may be associated with a fear of stigmatisation, an underappreciation of the benefits of these services or low parental literacy. These factors reinforce the inter-generational disadvantage described in Section 1.1.5.1. Social and child factors which increase the risk of poor language development may also put pre-schoolers at risk of poorer intervention response (Durkin & Conti-Ramsden, 2010).

Longfield (2019) estimate the average annual cost of speech and language therapy services to be £10.12 per child per year. Using an approximate 10% prevalence rate, this would suggest an average annual cost of speech and language therapy services to be around £100 per child with SLCN, although Longfield (2019) show that this is hugely variable by region.

Whilst it is not a necessary criterion for justifying interventions targeting childhood disorders, the return on investment for these interventions tends to far outweigh the initial cost. Whilst Law et al. (2012b) already suggest that interventions for SLCN are cost-effective, the methods used to estimate cost-effectiveness may underestimate the benefits of speech and language therapy.

1.3.1.2 Calculating return on investment: The need for life course data

Without much follow-up data, many studies of early childhood programmes report only a few early age outcomes like IQ or scores on a school readiness measure (Kline and Walters, 2016; RCSLT, 2016). The Stoke Speaks Out initiative, cited in a recent National Literacy Trust (2018) policy report, estimate a £1.19 return on investment per £1 spent by decreasing the percentage of children entering nursery with delayed language from 64% in 2002 to 39% in 2010. These findings are often extrapolated, as in Kline and Walters (2016) to predict longer term outcomes for the children receiving Head Start interventions, and in RCSLT (2016) where a higher return of £4.26 per £1 spent is estimated for the Stoke Speaks Out initiative once accounting for improved educational attainment, reduced risk of being Not in Education, Employment or Training (NEET) and a reduced risk of committing criminal offences.

Kline and Walters (2016) predict long term outcomes of the Head Start interventions in the United States by identifying children most likely to benefit from the programme and their expected change in education attainment as a result, given their background and current pre-school provision. Using this identification strategy and literature on the associations between education and income, they conclude that the long-term benefits of the programme would far outweigh the short-term gains. Whilst even the underestimated returns to speech and language therapy within the Stoke Speaks Out initiative suggest a positive return on investment, there is a clear undervaluation of childhood interventions when follow ups are limited to the benefits which materialise within the early years.

However, these projected extrapolations are limited by the predictive power of the model employed. Furthermore, they are limited by the areas measured, ignoring potential benefits or costs in other areas. For example, data on adult physical and mental health over the life course is not collected or projected, often due to the time scale of the project and the complexity of those predictions. However, health is crucial to both quality and longevity of life, and therefore life chances. This could cause a downward bias in cost benefit analyses as improved communication skills and education outcomes could potentially improve the health outcomes of the individuals invested in. Improvements in comprehension, for example, may improve the inequity of access to health services (Law et al., 2017b).

Within the more general field of childhood interventions, Barnett and Masse (2007) report a 2.5 benefit:cost ratio of an intervention which improved the education provided within free child-care provision. They account for projected benefits up to age 21, including improved projected earnings. Garcia et al. (2017) report a 7.3 benefit:cost ratio for the identical intervention. This is because Garcia et al. (2017) include data over an extra decade of life and a more comprehensive administrative data source for outcomes such as the cost of crimes committed by participants of the control and the treatment group.

Speech and language therapy interventions are very rarely evaluated beyond completion in a longitudinal design, such as that in Barnett and Masse (2007), which itself underestimated the true net benefits of the intervention by almost a factor of three. Longitudinal modelling is increasing the recognised positive value of early childhood interventions (Currie and Almond, 2011; Elango et al., 2016) and further analysis of outcomes into later adulthood could increase understanding of the potential impact on life chances of effective speech and language therapy.

In principle, projected modelling can deal with some of the issues raised. Marsh et al. (2010) project the difference in earnings for individuals receiving an effective language intervention, as described in Boyle et al. (2007), attempting to provide a cost: benefit analysis of speech-language therapy. However, this methodology is speculative. It is subject to underestimations of the sort proposed above, through ignoring potential benefits, and also subject to overestimations in cases where intervention effects may tail off after the intervention is removed (Falk et al., 2016). Within these cases, long-term gains may still be realised (Heckman et al. 2010; Heckman, Pinto, and Savelyev 2013) but little confidence can be attributed to extrapolative studies.

Similarly, retrospectively analysing a group of people with identified SLCN may not accurately represent the same group as those who would be identified prospectively. For example, the slow talkers in Reilly et al. (2010) would represent a very different population to any group who continue to experience communication difficulties into adulthood.

In summary, short follow ups commonly underestimate overall benefits, projected extrapolation cannot satisfactorily measure group outcome differences, and retrospective analysis cannot accurately model speech and language development. This thesis focuses on prospectively gathered data from birth into later adulthood in order to avoid these three common difficulties. Secondary data is selected which contains prospectively gathered information on childhood communication abilities as well as outcome data which covers a range of life outcomes necessary to capture life chances.

This thesis measures the long run outcomes for individuals with measured SLCN in childhood. This thesis does not model intervention effects for two key reasons: 1) data on the details of intervention received are limited in any data set; 2) no control group exists; and 3) intervention within the mid-20th century is unlikely to reflect modern intervention techniques which may often take place within a modern educational setting.

The reviewed literature does, however, highlight two important points. Firstly, that measuring the true impact of SLCN in childhood requires data extending beyond the early years. Secondly, that specialised assistance can be effective – if diagnosis is accurate (Hendricks et al., 2019) – that could mitigate the negative later life impacts of SLCN through service provision in early years.

1.3.2 Education and the Attainment Gap

Education can be considered a key vehicle for social mobility. The skills necessary to succeed in the academic school environment are hugely correlated with those necessary for success in the labour market (Heckman and Kautz, 2012). Current evidence suggests that children with SLCN are less likely to obtain educational qualifications at each level than typically developing peers, and that this differential is established by the first day of school.

1.3.2.1 From school entry

Within the UK, the Early Years Foundation Stage Profile results published annually by the Department of Education (with the exception of 2020 due to Covid-19), mark the percentage of children entering school with the necessary skills to access the school curriculum.

Communication and language form the first subset of the skills required for children to reach a 'good level of development' at the end of reception, aged five. This subset itself contains Listening and Attention, Understanding and Speaking. Data from the National Statistics (2018) demonstrate that only 28% of children with SLCN achieved a good level of development by age five, relative to 72% of all pupils. As SLCN at this age is unlikely to resolve (see Section 1.2.4, Stability of Speech and Language Development Profiles), these children are facing an initial deficit at school entry with the challenge to 'catch up' whilst their communication difficulties are likely to persist.

Language skills as early as two years old predict performance upon school entry (Roulstone et al., 2011). Once the child has started school, almost every aspect of their day will be impacted by their oral language skills, including their listening, attention, understanding and speaking. Participating in classroom activities, understanding instructions and socialising with peers are all dependent upon functional communication and this impacts school success (Foorman et al., 2012; Rubin et al., 2012).

Though communication is distinct from many other aspects of cognition, the ability to express ideas, spoken or on paper, is crucial for test scores, meaning language, communication, grammar, narrative skills and pragmatic skills are commonly classified as a subset of cognitive skills. Therefore, it is not surprising that children with communication difficulties tend to score worse on the sorts of standardised tests that have been typical in Education for decades, despite the potential for speech therapy to improve comprehension in children with SLCN (Lindsay et al., 2012).

Literacy scores are closely linked to earlier oral language ability, with persistence demonstrated into adulthood (Law et al., 2009). Snow (2020) outlines a 'language house' theory of reading development, depicting early oral language skills as the foundation upon which subsequent reading and interpersonal skills are built, leading to greater employment, educational and social engagement outcomes in adulthood.

The reference made to interpersonal skills, the pragmatic skills for communication related with forming bonds with peers, are shown to impact classroom participation (McClelland et al., 2007). Whilst the relationship between language ability and literacy, particularly reading, may be intuitive, less language-focused subjects such as maths also exhibit high levels of disability within SLCN groups (Komesidou and Hogan, 2019). Whether this is due to general cognition, co-morbid factors, home environments, understanding of educators' instructions, or classroom participation is unclear, but these results are well replicated. McLeod et al. (2019) similarly demonstrate within an Australian cohort that SLCN from age four to seven predicts lower literacy and numeracy scores from ages eight to 12.

Structural equation modelling has demonstrated that the transmission of socio-economic advantage, as measured by parental income predicting academic skills at age 10, are mostly mediated by the income gap in vocabulary skills observable at age four in a large US sample (Durham et al., 2007). The importance of early vocabulary in predicting academic achievement is clearly relevant to questions of social mobility and inequality.

1.3.2.2 Academic Qualifications

Within the UK, GCSEs are the most common form of qualification upon leaving compulsory schooling. Literacy and numeracy, represented by English and Maths GCSEs, are commonly considered the gateway qualifications for future education and employment and are present in most national statistics. National Statistics (2019) demonstrate that, in 2018, 21% of pupils with SLCN as a primary special educational need achieved pass grades (C or above at the time, now 4 or above) in English and Maths GCSE compared to 64% of all pupils. Put simply, the achievement gap observed at school entry between SLCN pupils and the general population does not narrow by the end of compulsory schooling. The importance of this risk factor is consistent from early years into adolescence for predicting school attainment.

A crucial element of literature here draws back to Section 1.2.4, The Stability of Speech and Language Development Profiles. Snowling et al. (2001) compare the GCSE performance of four groups – 1) persistent specific language impairment (n=24, now known as Developmental Language Disorder); 2) language impairment resolved by age 5;6 (n=22); 3) children with general delays, including language impairment (n=14); and 4) age-matched controls (n=43). The GCSE outcomes were generally poorer for the children with language impairments, regardless of general delays in other areas. Continuation into full-time education was less common amongst the SLCN groups relative to the age-matched controls. However, for the resolving group, overall results were much closer to the age-matched controls.

The resolving group in Snowling et al. (2001) entered around the same number of GCSE examinations, although they achieved around two A-C grades fewer on average. 41% of the resolved specific language impairment group achieved 5 A-C grades, relative to the national average of 47.8%.

The persistent group, by comparison, achieved fewer than 1 A-C grade on average, despite sitting approximately the same number of examinations as the resolving and the control group. There was considerable heterogeneity, as is a common finding throughout the literature (Conti-Ramsden et al., 2018), such that many children with resolving or persistent language disorders go on to perform exceptionally well at GCSE and beyond.

More recent studies have suggested that outcomes at age 16 do not necessarily limit the overall qualification levels attained by individuals with language disorders. Durkin et al. (2009) report that 91% of children with Developmental Language Disorder (DLD), the persistent form of language disorder, continued to engage in either further academic education or some form of training beyond compulsory education, at age 17, within their sample. However, the attainment gap observed at age 16 in Durkin et al. (2009) supports the general findings of Snowling et al. (2001), that early communication difficulties suggest a significantly lower chance of achieving 5 A-C grades at age 16. Durkin et al. (2009) also find that the lower performance of the DLD group is not reflection in their satisfaction rates, suggesting that initial expectations are likely to have been lower.

Attainment at GCSE is generally predictive of both attendance and performance within further education (Crawford and Greaves, 2015). Further education attendance and performance within children with SLCN was unmeasured at the commencement of this thesis. Conti-Ramsden et al. (2018) have since confirmed that this finding is similarly true for this population. The sample they analyse is drawn from the Manchester Language Study, as in Durkin et al. (2009), who had reached age 25 by the time of analysis.

26% of the group with Developmental Language Disorder did not achieve even D-G grade GCSEs, relative to 0% of the age-matched control group. 18% of the DLD group achieved A-Levels, relative to 72% of the age-matched control group. And finally, only 10% of the DLD group achieved a university degree, relative to 41% of the age-matched control group. Achievement of some formal academic qualification, A-Levels and a degree were considered to be the three key milestones in educational achievement within this UK study (Conti-Ramsden et al., 2018) and the children who experienced severe language disorders were significantly less likely to reach each of these milestones.

1.3.2.3 Changes to the UK Education System

Conti-Ramsden et al. (2018) studied the key milestones in the modern UK Education System. However, this system has undergone many substantial changes over the course of the 20th century which are relevant to this thesis.

The school minimum leaving age was changed from 15 to 16 in 1972 (Bolton, 2012), meaning there are slight differences across the cohorts analysed within this thesis. Exogenous policy increases in compulsory years of education have been causally linked to increases in life expectancy, partially through certain health behaviours (Brunello et al., 2015). A higher minimum level of education is likely to improve the outcomes for people with SLCN on average given the high proportion of this group who leave school at the minimum compulsory age (Conti-Ramsden, 2018).

The largest change in education to occur in the UK within the 20th century was the move from the tripartite system of secondary education, introduced in the Education Act of 1944, to non-selective secondary education encouraged by the 1965 circular 10/65. Comprehensive schools became the most common form of secondary school where they had previously been a rare experiment (Bolton, 2012). This means that most children in the first cohort studied (Newcastle Thousand Families Study, 1947) experienced selective secondary education, being sorted into Grammar, Technical, or Secondary Modern schools based upon test scores at age 11. Most children in the second cohort studied (British Cohort Study, 1970) did not.

Very few children attended technical schools nationally, and the age 11 exams primarily separated children into Grammar schools or Secondary Moderns based upon local pass/fail cut-off grades (Bolton, 2012). For the children who struggled with literacy within primary school, segregation according to ability at age 11 could have a number of effects. Given the attainment gap persisting from school entry, described in the previous section, and the lower numeracy and literacy scores, children with SLCN may be more likely to have attended Secondary Modern schools. If this tendency increases on a national scale, it could have a number of impacts on the outcomes for children with SLCN at the end of compulsory education and into employment.

Firstly, segregation may decrease the exposure to lexical diversity, potentially impacting vocabulary development. Other peer group effects may occur which are likely to be negative for the children attending schools without selective admission scores. Secondly, children at

Secondary Moderns were more likely to be entered for CSE examinations, as opposed to O-Levels, which were considered to be less academic (Bolton, 2012). Thirdly, Durkin et al. (2009) demonstrate that self-reported satisfaction rates of equivalent grades are higher within individuals with DLD in a modern cohort. This may reflect a lower level of self-expectation amongst individuals with DLD and a selective schooling process may reinforce negative self-perceptions. Finally, an earlier measurement of 'ability' allows less time for children to address initial communication difficulties and to catch up to their peers. In cases where SLCN would have resolved within adolescence, however rare, a key milestone for future achievement would pre-date this resolution.

Most children now attend mainstream comprehensive schools. The Department for Education (2016) report that educational opportunities have improved for young people with disabilities for children born closer to the end of the 20th century. Tests at age 16 are known as GCSEs, replacing the O-level system since 1988. Despite a recent change to grading, the achievement of five pass grades (scores from 9-4 being broadly equivalent to grades A*-C) including Mathematics and English still represents a key milestone in Key Stage Four achievement. The number of children achieving this milestone has accelerated within the previous 30 years, particularly for females (Bolton 2012).

Cocking and Mestre (1988) claimed that "language is the currency of education". They stated that language skills in childhood were valuable in the development of broader cognitive and social skills necessary for success in life. The subsequent 32 years of research have provided strong evidence for this argument.

The education landscape has changed dramatically over the course of the 20th century and beyond. Research extending from the 1990s (Records et al. 1992) to the past few years (Conti-Ramsden et al., 2018) consistently suggest that the early attainment gap observed for individuals with SLCN persists throughout academic education. This lived experience of individuals – contributory to individual, parental and societal perceptions of the role of education for future opportunities – is of interest to the central questions of this thesis: What does life look like for a child with early communication difficulties? How and why are their lives different from children with typical speech and language development?

1.3.3 Non-cognitive skills

This thesis seeks to improve understanding of the impact of early communication difficulties over the life course. The ability to verbally communicate is a skill which is developed from infancy and continually used throughout life. The collection of all skills, knowledge and experience controlled by individual are commonly referred to as their 'human capital', measured by their value to the individual over the life course. This value is derived from the opportunities available to an individual, and their success in making the most of those opportunities. This thesis draws upon the literature regarding early human capital development and particularly the work of Nobel Prize-winning economist James Heckman. Heckman and Kautz (2012) emphasise two essential points which are relevant to the current thesis:

1. Human capital formations are endogenous; endowments and investments accumulate over time. This increases the relative importance of differences that emerge early in life.
2. Non-cognitive skills, relative to cognitive skills, have been historically under-researched and under-valued in educational settings. These skills have been shown to be equally and incrementally valid predictors of later life outcomes.

The first point echoes the findings of Section 1.2.4 of this thesis relating to the development of language over the life course. Vocabulary at any age is predicted by vocabulary at an earlier age, and the value of investments – such as the home learning environment, exposure to a range of vocabulary from parents, increased frequency of parental or independent reading – are recognised throughout the literature reviewed throughout Section 1.2 of this thesis. The notion that early development has long-lasting impact over the life course is the central concept underpinning this thesis.

The second point is reflected in the emergence of research relating SLCN to non-cognitive skills within the past decade. For years, researchers within speech and language sciences have recognised the need to control for and quantify children's more general cognitive development (Morley, 1965) whereas the associations between early language development and non-cognitive development have only recently gained direct attention more recently (Ullrich et al., 2017; Toseeb et al., 2017).

This section of the literature review will first outline the terminological approach to non-cognitive research before synthesising the key evidence pertaining to the importance of non-cognitive skills. Finally, the limited evidence relating speech and language development to the development of non-cognitive skills in childhood and adolescence are summarised.

1.3.3.1 Terminology

Non-cognitive skills have also been known by "character", "personality", "soft skills" and "socio-emotional skills". Some studies draw a distinction between non-cognitive 'traits', often perceived to be static, and non-cognitive 'skills', which suggests malleability. This thesis will typically use the term 'skills', due to the potential for change and gradual formation (Cunha et al., 2010), particularly when referring to childhood or adolescence. 'Traits' may be more commonly used to refer to measures taken in adulthood, implying stability (Barrick and Mount, 1991). However, as this thesis does not measure the impacts of any interventions, non-cognitive or otherwise, this distinction is not a focus of this research and the terms can be interpreted interchangeably, as in much of the reviewed literature (Almlund et al., 2011; Heckman and Kautz, 2012).

Heckman and Kautz (2012) define these skills as "personality traits, goals, motivations and preferences that are valued in the labour market, schools and many other domains". These skills revolve around dealing with others as well as mastering and motivating one's own behaviour (Goodman et al., 2015).

1.3.3.2 Taxonomies of Non-Cognitive Skills: The OCEAN Characteristics

Within personality psychology, the importance of certain traits has been explored for almost a century (Terman et al. 1925). Personality traits, referred to as non-cognitive skills, have more recently begun to gain acceptance across a wide variety of fields as potentially causal producers and predictors of economic performance and personal wellbeing (Heckman and Kautz, 2012).

McCrae and Costa (1992) developed the 'Big 5', or the 'Five Factor Model'. These five non-cognitive skills are intended to represent the latitude and longitude of personality and are the most widely used parameterisation of non-cognitive skills within the literature (Almlund et al., 2011; Heckman and Kautz, 2012). McCrae and Costa defined these traits as below:

- Openness to Experience: The tendency to be open to new aesthetic, cultural or intellectual experiences.
- Conscientiousness: The tendency to be organised, responsible and hardworking.
- Extraversion: Orientation of one's interests and energies outwards. Characterised by positive affect and sociability.
- Agreeableness: The tendency to act in a cooperative, unselfish manner.
- Neuroticism/Emotional Stability: (Low/High) Predictability and consistency in emotional reactions and (High/Low) susceptibility to rapid mood changes and psychological distress". Emotional stability contains an important sub-skill, commonly measured within longitudinal studies and therefore of relevance to this thesis, as defined by Almlund et al. (2011):
    - Locus of Control – Internal or external. An individual with an internal locus of control believes that life events are typically caused by their own actions.

These five factors, commonly referred to as the "OCEAN" characteristics (an acronym of the skills, although this thesis focuses on emotional stability rather than neuroticism as the intention is to measure positive valuable skills), develop early in life. Whilst heritability plays a role, these skills can be augmented and enhanced through guidance and instruction (Heckman et al., 2013). Quality parenting and caring environments are equally important for non-cognitive skills as cognitive skills (Cunha et al., 2010), as is evident in the development of speech and language.

The key interest in non-cognitive skills stems from their malleability in adolescence, with stability increasing with age (Heckman and Kautz, 2012; Roberts and Wood, 2006). Whilst cognitive ability has high rank stability from age 12, non-cognitive skills can still be targeted throughout adolescence (Walsh, 2005). Parenting, education and environment have all been targeted within non-cognitive interventions in adolescence (Heckman and Kautz, 2012) whilst most interventions in speech and language therapy are targeted at the same delivery agents but within primary school age children (Dockrell et al., 2014; Lindsay et al., 2012). The targeting of the context around a child as the mechanism for change is important in expanding the influence of the intervention beyond the allocated time in which it is carried out. Shonkoff and Phillips (2000) state that speech and language interventions should target the most malleable factors related to the recipient to maximise the impact on childhood trajectories.

This suggests a potential important factor in the life outcomes of children with SLCN with direct clinical relevance.

Cunha et al. (2010) emphasise the importance of investments in improving non-cognitive skills, alongside early endowments, for enhancing non-cognitive skills present throughout adulthood. Kautz et al. (2015) summarise the essential ingredients for promoting non-cognitive skills as follows:

1. A powerful role for families and social environments.
2. A recognition of the skills required for economic comparative advantage, as well as personal satisfaction.
3. An understanding of how skills are formed. Initial skills interact with investments to develop further skills (see Cunha et al. (2010) or Heckman et al. (2013) for structural longitudinal analyses).

When reviewing the literature relating to non-cognitive skills, it is immediately clear that many of the fundamental processes for development of these skills resonate with the literature within speech and language sciences; the importance of families and social environments (Shonkoff and Phillips, 2000; Hart and Risley, 1995; Law et al., 2010); the setting of personal goals within intervention (Dockrell et al., 2014); the role of initial development alongside intervention (Reilly et al., 2010; McKean et al., 2017b; Law et al., 2017c; Dockrell et al., 2014).

Social gradients are a running theme within most social science research. OCEAN characteristics are related to parental advantage through similar mechanisms to those discussed in Section 1.2.5.1 – family resources, environmental influences, parental input and heritable traits all play some role in the formation of non-cognitive skills. Boneva and Rauh (2015) hypothesise that parental perceptions of the value of educational investments helps to explain these social gradients. Parents from low-income households tended to attribute relatively higher values to educational investments made later in childhood. Within the context of limited resources, educational investments at different time points may be more likely to be treated as substitutes. The endogenous nature of human capital formation suggests that investments at early time points are the most valuable. This suggests a market inefficiency due to a lack of information which could be addressed through policies and schemes targeted within the early years.

Findings from Fletcher and Wolfe (2016) demonstrate the importance of these approaches for tackling social inequality. They find that non-cognitive skills (self-control, co-operation, impulsivity-restraint, and aggression) diverge over adolescence, and that this divergence is predicted by family income. This also suggests that later measures of non-cognitive skills may emphasise social gradients in outcomes, as has been found for cognitive skills (Case et al., 2002), as the divergence in skills across the socio-economic spectrum is widest at later time points. Non-cognitive skills measured from as young as five are predictive of economic outcomes (Blanden, 2007; Fletcher and Wolfe, 2016) and childhood trajectories may be best targeted before social gradients increase, as with language development (Taylor et al., 2013).

Despite clear parallels between the human capital literature focusing on the OCEAN characteristics and the wealth of research into detailed development of communication skills within speech and language sciences, very few studies measure the relationship between these skills either concurrently (see van Agt et al., 2010 for an exception) or longitudinally through development (see Toseeb and St Clair (2020) for an exception regarding "prosociality", a subset of 'Agreeableness'). To the author's knowledge, no studies have measured the non-cognitive skills in adulthood of individuals who experienced SLCN in childhood.

The following section summarises the value of many non-cognitive skills in adulthood. The papers selected for this review contain at least one outcome directly measured in the later analyses. This clarifies the importance of the traits tested in the analysis chapters of this thesis for life course research.

1.3.3.3 The importance of non-cognitive skills

Heckman and Kautz (2012) summarise a significant body of evidence demonstrating that educational attainment, health and labour market success are predicted by non-cognitive skills at a similar are linked strongly by more than just cognitive ability. Non-cognitive skills are shown to be similarly valuable in these economic outcome measures to cognitive skills.

Cunha et al. (2006), for example, find that 16% of variation in educational attainment can be explained by adolescent cognitive traits, whilst 12% of the variation is additionally explained by adolescent personality. The finding that cognitive and non-cognitive abilities are complementary in their explanation of variance is persistent throughout the literature (Heckman and Kautz, 2012). Heckman et al. (2013) demonstrate that the omission of one of these forms of ability will result in over-estimation of the importance of the other due their co-linearity.

Cunha et al. (2006) find a further 15% of the variation in educational attainment (supplementary to the 16% and 12% explained by cognitive and non-cognitive traits respectively) could be explained by measured parental investments. Heckman and Kautz (2012) summarise a number of papers which support the finding that cognitive and non-cognitive skills can be developed through investments over time with significant long-term payoffs, hence the use of the phrase 'human capital'. The potential for positive intervention effects, whether intentionally targeting non-cognitive skills or not (see the analysis of the Perry Pre-School Programme within Heckman and Kautz (2012) for an example of incidental impact arising from non-cognitive skill enhancement), provokes the interest in non-cognitive skills from a clinical standpoint.

The OCEAN characteristics are shown within Heckman and Kautz (2012) to collectively predict long-term outcomes, although which specific characteristic – and which trait within the overall characteristic – predicts which outcome can vary. Amongst the "Big 5" OCEAN characteristics (McCrae and Costa, 1987), 'Conscientiousness', the trait relating to work ethic and organisation, is generally seen as the most strongly predictive of labour market, health and educational outcomes. It ranks as strongly as measures of cognitive ability and has been shown as the best compensator for lower cognitive ability (Roberts et al., 2007; Almlund et al., 2011; Borghans et al., 2008).

Conscientiousness has been shown to predict job performance, albeit less so than intelligence, a cognitive trait. Work ethic and attention to detail, skills contained within conscientiousness, are clearly transferable across a number of jobs. However, conscientiousness has also been found to predict job performance more so than intelligence within less complex jobs (Heckman et al., 2013; Heckman and Kautz, 2012). Whilst all OCEAN characteristics have been found to predict some health outcomes, conscientiousness is again the most predictive (Kautz et al., 2015). Conscientiousness is predictive of life expectancy, through pathways of education and health behaviours, to a greater extent than socio-economic status, IQ or any other OCEAN characteristic (Roberts et al., 2007).

Openness to experience is the personality trait most closely related to education. Enjoyment of educational experiences will encourage learning and increase the probability of choosing to engage in further education. Exposure to educational experiences, such as through attendance at a further education institution, can enhance and individual's enjoyment of educational experiences (Heckman and Kautz, 2012). Openness to experience has been found to be more predictive of outcomes than Conscientiousness amongst groups of low socio-economic status (Falk et al., 2016). These two factors make this trait potentially highly valuable to individuals with SLCN who may have negative experiences of schooling through bullying (Byers et al., 2012) or difficulties with peer relationships (Conti-Ramsden et al., 2013) and are substantially more likely to come from a disadvantaged socio-economic background (Law et al., 2013). Sections 1.3.2 and 1.2.5.1 cover the importance of education and SES for children with SLCN in more detail.

High levels of openness to experience is beneficial to many economic outcomes, including employment and earnings, as well as health outcomes (Heckman and Kautz, 2012). However, controlling for education tends to significantly reduce the importance of this trait (Heckman et al., 2010; 2011). This may explain the trait's relatively high value within low SES groups, where education can often be seen as the vehicle for social mobility.

Emotional Stability, often parameterised in reverse as 'Neuroticism', similarly predicts specific labour market outcomes. The amount of effort an unemployed person puts into searching for a new job is significantly correlated with high levels of emotional stability (Kautz et al., 2015). High levels of self-esteem (an individual's subjective estimation of their own self-worth, Almlund et al. (2011)) and an internalised locus of control are highlighted as potentially causal producers of this increased search effort and subsequently lower rates of long-term unemployment. Heckman et al. (2006) use a sequential model to demonstrate that self-esteem and locus of control within adolescence predict earnings in adulthood to a stronger degree than cognitive measures in adolescence, particularly for individuals who followed atypical education trajectories (high school drop-outs, for example).

Within the British Cohort Study (1970), measures of conscientiousness, self-esteem and locus of control are taken at age 10 and 16 from maternal likert-scale ratings. Prevoo and Ter Weel (2015) demonstrate that these measures predict employment, education and health behaviours up to age 38. Blanden (2007) previously utilised similar measures within this cohort and finds that non-cognitive skills from as young as five are predictive of earnings up to age 30. There is opportunity to improve non-cognitive skills throughout childhood, yet skills measured within the first years of life are significant predictors of outcomes into adulthood within population data. A similar frustration of a perceived missed opportunity to improve lives is sensed within studies of early language development (Law et al., 2017a).

The relative impact of non-cognitive skills and cognitive skills are heterogenous, partially dependent on the outcome. Lindqvist and Vestman (2011) and Heckman et al. (2013) both find that non-cognitive skills are most predictive of unemployment and earnings within low-wage samples, whereas cognitive skills are most predictive of wages for high earners. These papers use very different samples, with one drawn from a Swedish military exam and the latter drawn from an American large longitudinal dataset. Conti et al. (2010; 2011), by contrast, find that employment and wages are most predicted by cognitive ability, whereas health behaviours are more predicted by non-cognitive skills.

Goodman et al. (2015) compare the impacts of non-cognitive and cognitive skills, measured at age 10, for predicting outcomes to age 42. Whilst cognitive skills were more correlated with income, social skills were more correlated with self-reported happiness. This emphasises the feeling that a broader range of outcomes can highlight important differences in life experiences, particularly given the broad focus of parents discussing the outcomes they perceived to be important for their children (Dockrell et al., 2014; Great Ormond Street Hospital, 2018).

Analysis of the life experiences of individuals is therefore likely to be incomplete when non-cognitive skills are not accounted for. Non-cognitive skills play a significant role in determining many of the life outcomes of interest to this thesis, including those previously shown to be typically poorer for individuals with SLCN. A child’s ability to communicate may be essential to developing and applying these non-cognitive skills although this remains untested.

The following section summarises the current knowledge relating speech and language development to non-cognitive skills. The papers reviewed contain at least one facet of the OCEAN characteristics, enabling the development of hypotheses for the later analysis chapters.

1.3.3.4 SLCN and Non-Cognitive Skills

Socio-Emotional and Behavioural (SEB) difficulties have been linked to early speech and language acquisition for decades (see Beitchman et al., 1996 for a review). Behaviour itself is a form of communication and externalised behavioural issues are unsurprisingly more common amongst children who struggle to communicate verbally. The relevance of these behavioural difficulties to academic attainment has also been noted in the SLCN literature (Dockrell and Lindsay, 2000; Dockrell et al., 2007). However, there is a paucity of research investigating the positive development of non-cognitive skills for individuals with SLCN directly, despite most social competency measures including interpersonal communication skills (Gallagher, 1993). Furthermore, many of the noted correlations between SEB difficulties and SLCN are established within clinical or high-risk samples, as opposed to the general population (Rose et al., 2018).

Almond et al. (2018) review the impact of early human capital formations for predicting later life outcomes. Within measures of early human capital formation, Almond et al. report the use of the Peabody Picture Vocabulary Test (Dunn and Dunn, 1997), a measure of receptive vocabulary, in many models. In these human capital models, the development of communication is considered complementary to the development of both cognitive and non-cognitive skills.

The possible impact of early communication development on early non-cognitive skills is clear from the literature. The works of Heckman, Kautz and Cunha discussed in the prior section show that early non-cognitive skills are strongly related to later non-cognitive skills, although there is huge potential to influence the development of these skills through similar mechanisms to those influencing communication development (Heckman and Kautz, 2012; Lindsay et al., 2012a; Rose et al., 2018).

The implications of this issue from the perspective of education policy are considerable. In 2016, the Education Endowment Foundation (2016) put out a call for interventions that promote “character” – a proxy for the Big 5 non-cognitive skills – before the government provided funding for similar interventions (Gov.uk, 2016). Current Department for Education (2019) policy contains framework guidance for ‘Character Education’, including specific reference to traits including self-control, self-efficacy and social confidence, targeted at schools. Damian Hinds in his first speech as Education Secretary encouraged schools to explicitly focus on soft skills to better prepare children for the labour market (The Guardian, 2018).

Furthermore, parents of children with speech, language and communication needs have directly stated a desire to see their children develop independence and build social networks (Roulstone et al., 2012). This follows a large amount of research detailing the poorer peer relationships (Conti-Ramsden et al., 2013) and school experiences (Byers et al., 2012) of children with SLCN. This implies that parents are concerned about their children's development of certain non-cognitive skills in childhood and adolescence.

Rose et al. (2018) demonstrate that early socio-emotional regulation, from age three to eight, is fostered through a positive home learning environment (with specific reference to literacy) and language acquisition (both receptive and expressive) at age three. Children with the strongest language skills demonstrate the highest levels of co-operative behaviour and emotional self-regulation, whilst demonstrating the lowest levels of aggressive behaviour.

The links between emotional stability and childhood social-emotional or behavioural difficulties are intuitively relevant and evidently present for children with communication difficulties (Lindsay and Dockrell, 2012; Benner, 2002). In fact, behaviour problems in school often lead to identification of Speech, Language and Communication Needs that had been missed earlier in life (Beard, 2018). Early temper tantrums and behavioural problems are related to communication development, theoretically through the frustration a child experiences when they are unable to get their needs across to the primary caregiver (Shonkoff and Philips, 2000). This has been linked to youth offending with teenage populations and their tendency towards poorer speech and language skills (Snow and Powell, 2012).

Table 1.2 synthesises the body of research relating the OCEAN characteristics to speech and language development. Most research is focused on negative traits noted in early childhood for individuals with identified SLCN. There is evidence that these negative associations may persist, or even worsen, over time (Botting and Conti-Ramsden, 2000; Conti-Ramsden and Botting, 2004; Voci et al., 2006) although little research has focused on the fostering of positive skills (see Toseeb et al., 2020 for a recent example of the movement towards this).

Table 1-2 Synthesis of Big non-cognitive, big five/OCEAN-related research in speech and language sciences

Note: Facets attributed to OCEAN characteristic using John and Sririvista (1999), Roberts et al. (2007) and Heckman and Kautz (2012). Facets are grouped where they have been commonly grouped within research papers.

* denotes a trait contained within more than one OCEAN characteristic.

OCEAN Factor	• **Facets / Traits** – associations with SLCN within literature search
Openness to Experience	• **Sensory sensitivity** – children with DLD show no differences in auditory sensitivity, although lower rates of auditory filtering and lower general sensation seeking (Simpson et al., 2020). • **Wide actions and interests** – lower engagement in further education (Conti-Ramsden et al., 2018). • **Pleasure in low-intensity activities / Curiosity / Imaginative / Artistic / Excitable / Ideas / Unconventional values** – No notable research
Conscientiousness	• **Attention / Concentration / Impulse control** – lower levels of attention span and control of impulses amongst individuals with language impairment (Levickis et al., 2018b; Javorsky, 1996). Increased rates of behavioural disturbance from age five to 12 (Beitchman et al., 1996). Higher levels of hyperactivity demonstrated within young children with language impairment (Levickis et al., 2018b; Clegg et al., 2015). Hyperactivity and inattention are also related to poorer vocabulary growth between the ages of four and nine (Westrupp et al., 2020). Self-control in children with early language difficulties predict later literacy and maths skills (McClelland et al., 2007). • **Ambitious** – equal self-reported satisfaction levels despite poorer education outcomes (Durkin et al., 2009). • **Persistence** – lower amongst with speech or language impaired children aged five (Harrison and McLeod, 2010). • **Effort control / Self-discipline / Activity*** – No notable research.

Extraversion	• **Social Dominance** – lower social dominance (higher levels of social withdrawal, the opposite of social dominance) within individuals with SLCN (Lindsay and Dockrell, 2000; van Daal et al., 2007; Fujiki et al., 2002), particularly for girls with expressive language difficulties (Beitchman et al., 1996). • **Shyness*** - Higher levels of shyness (Wadman et al., 2008; Durkin et al., 2017). High rates of social phobia (Beitchman et al., 2001) into adolescence (Voci et al., 2006). • **Self-confidence** – lower self-confidence amongst individuals with SLCN (Lindsay and Dockrell, 2000; Fujiki et al., 2002), particularly relating to social confidence and shyness (Durkin et al., 2017). • **Sensation seeking / Vitality** – below the typical range for school-age children with DLD (Simpson et al., 2020). • **Activity* / Positive emotionality / Sociability / Warmth** – No notable research
Agreeableness	• **Aggressiveness / Compliance** – higher rates of aggression in adolescence for boys with childhood SLCN (Brownlie et al., 2004), particularly language difficulties (van Daal et al., 2007). Poor language in adolescence, regardless of diagnosis, is common in youth offending institutions (Snow and Powell, 2012). Slower decline in conduct problems for language impaired low SES boys than for typically developing children (Yew and O'Kearney, 2015). • **Sympathetic / empathetic / trusting / co-operative** – poorer peer relationships and interactions (Valance et al., 1998; Conti-Ramsden et al., 2013; Manso and Garcia, 2005; Durkin and Conti-Ramsden, 2010; Clegg et al., 2015), including peer rejection and bullying (Byers et al., 2012; Menting et al., 2011). These peer problems mediate some of the relationships between language and emotional problems (Forrest et al., 2018). Lower co-operation at age eight related to expressive and receptive language at age three (Rose et al., 2018) and con-currently at age eight (van Agt et al., 2010).

	• **Prosociality** – Mixed evidence. Girard et al. (2015) find lower levels of prosociality amongst language-impaired individuals, although most studies find this to be an area of relative strength, with scores within the normal range throughout adolescence and into adulthood (Toseeb et al., 2020; 2017; Toseeb and St Clair, 2020; Wadman et al., 2008; Durkin and Conti-Ramsden, 2010). • **Wilfulness / Trusting / Irritability*** – No notable research
Emotional Stability	• **Inhibition / Fearfulness / Shyness*** - Higher levels of shyness (Wadman et al., 2008; Durkin et al., 2017) and social phobia (Beitchman et al., 2001) into adolescence (Voci et al., 2006) for children with SLCN. However, van Agt et al. (2010) find no difference in 'worried behaviours' between children with and without language disorders. • **Frustration** – Expressions of temper tantrums are common in children with poor verbal communication (Shonkoff and Phillips, 2000). • **Anxiety** – Higher rates of anxiety amongst individuals with poor language (Law et al., 2009; Durkin and Conti-Ramsden, 2010) into adulthood, particularly social anxiety (Beitchman et al., 2001). • **Internal Locus of Control** – less internalised locus of control represented by reduced independence in decision making and task completion (Markham et al., 2009). Lower mental health, including a sub-measure of internal locus of control, into adulthood (Law et al., 2009). • **Self-esteem** – Lower self-esteem established into adolescence for individuals with SLCN (Wadman et al., 2008; Schoon et al., 2010a; Lindsay and Dockrell, 2000; Fujiki et al., 2002). • **Self-efficacy** – lower self-efficacy shown into adulthood (Botting et al., 2016; Durkin et al., 2017), supported by Law et al. (2009) using self-efficacy within a mental health measure. • **Optimism / Depression / Irritability* / Sadness** – No notable research.

The most consistent finding relates to the child's experience of social interactions throughout the school years and into early adulthood. Conti-Ramsden et al. (2013) find that adolescents with language impairment report around 12 times as many peer problems as typically developing peers. The finding that children who struggle to communicate also struggle to develop positive social confidence and peer relationships is consistently found across the literature (Conti-Ramsden et al., 2013; Wadman et al., 2008; Durkin and Conti-Ramsden, 2010; Byers et al., 2012; Menting et al., 2011; Durkin et al., 2017; Beitchman et al., 2001). However, there is a similarly consistent finding that this is not due to a lack of willingness (Toseeb et al., 2017; Durkin and Conti-Ramsden, 2010; Wadman et al., 2008), aside from through a fear of further negative experiences (Beitchman et al., 2001; Voci et al., 2006). Toseeb et al. (2020), for example, find that "play" in early childhood can foster the skill of prosociality amongst individuals with DLD, which in turn can protect against later externalised difficulties. Activities, such as "play", may help to foster an enjoyment, rather than a fear, of new social experiences and the current literature suggests that these activities should be encouraged for helping to develop crucial non-cognitive skills. This notion appears to be supported by the views of parents of children with speech, language and communication needs (Dockrell et al., 2014).

Whether encouraging social interactions in early childhood could similarly protect against the negative mental health outcomes associated with social anxiety remains to be tested. Furthermore, whether early negative experiences socially in school may help to shape the individual's view of schooling is unknown. This could influence the observed lower rates of later academic attainment and engagement through an individual's openness to experience, particularly in families of low socio-economic status. The targeting of non-cognitive skills is an area with huge potential benefits for individuals with SLCN although the current literature is insufficient to justify any adjustment of policy.

No papers have explicitly tested whether the Big 5 non-cognitive skills are associated with speech and language development in childhood. Van Agt et al. (2010) report findings relating to extraversion, agreeableness, and emotional stability, although each of their measures are limited to one facet of the broader characteristics and reported as such in Table 1.2. Through synthesising the available research and relating them back to the underlying traits within each broader characteristic, we build the following hypotheses. Individuals with SLCN, relative to typically developing peers, may develop:

- Similar levels of openness to experience. Very little research has investigated these traits, although negative experiences of socialising in school may theoretically limit the development of openness to experience for individuals with SLCN.
- Similar or lower levels of conscientiousness. The only noted associations in the literature stem from attention span in young children, with very little research regarding other traits of conscientiousness.
- Lower levels of extraversion. This effect is hypothesised from a wide range of papers at various ages, samples, and underlying traits.
- Lower levels of some agreeableness traits, but not prosociality. This effect is hypothesised from low levels of co-operation and peer relationships in multiple studies, although self-reported measures may emphasise internal willingness to communicate (prosociality) more so than teacher-reported co-operative behaviours.
- Lower levels of emotional stability. This effect is hypothesised from the noted social anxiety and shyness within SLCN groups and may not apply to other traits of emotional stability.

A theme throughout the reviewed literature is that there is considerable heterogeneity in groups of people with speech, language, and communication needs. Whilst many research papers observe group-level mean differences in outcomes, this does not suggest that all individuals with SLCN will struggle in any given domain. The same is true for non-cognitive skills. Durkin and Conti-Ramsden (2010), as well as Mok et al. (2014), emphasise that many children with speech or language difficulties will develop non-cognitive skills as well as any typically developing child. Similarly, a child with excellent verbal communication skills is not immune to having low levels of any OCEAN characteristic. What is clear is that more research should be done to establish which non-cognitive characteristics are impacted by early speech and language development, and whether these characteristics can be targeted in order to improve the lives of individuals with early communication difficulties.

1.3.4 Health

Speech, language, and communication needs are a highly prevalent childhood disability. Speech and language therapists are healthcare professionals who seek to improve their patient's communication skills. As with other healthcare professionals, the overall aim is to improve the well-being of the individual. Health represents one essential form of the quality of life experienced by an individual.

The early ability to communicate is unlikely to be the only form of health difficulty experienced by individuals with SLCN, as research shows many correlated health problems at various ages. Hearing loss and neurodevelopmental difficulties (e.g., Attention Deficit/Hyperactivity Disorder/ADHD) may impact the child's development of speech and language (Law et al., 2017a; Harrison and McLeod, 2010) through a reduced accessibility to language exposure in the home and school environments. Whilst most children with language difficulties experience them in isolation (Norbury et al., 2016), for a minority they may occur alongside other child health disabilities which are not caused by the speech, language or communication needs of the child.

1.3.4.1 SLCN and health

Decades of research have indicated the heightened risk of psychiatric disorders amongst children with SLCN (see Beitchman et al., 1996 for an early review of this literature), with many referrals to psychiatric units having previously undiagnosed SLCN (Cohen, 1996). As children struggle to communicate, one externalised form of communication employed may be behaviour, and poor behaviour is extremely common in children with SLCN, extending into adolescence (Snow and Powell, 2012; Ripley and Yuill, 2005). Not all frustrations are externalised, however, with internalised difficulties similarly demonstrating an enduring connection with SLCN throughout childhood and adolescence (Conti-Ramsden et al., 2013).

There are differences within SLCN groups in terms of health outcomes. Lack of improvement despite intervention and the SLCN being primarily receptive are particularly associated with poor prognosis (Snowling et al., 2006; Beitchman et al., 1996).

Le et al. (2020a) report that children with low language, as measured from age four to 11, are significantly less likely than children with typical language to follow stable trajectories at high levels of health-related quality of life (HRQoL), as measured from age four to 13.

Approximately a third of children with low language experienced low trajectories of HRQoL with a rapid decline over time, and a further 26% follow a trajectory of reduced HRQoL with a slow decline over time. Within the same cohort, McKean et al. (2017b) had previously demonstrated that low language co-occurs with HRQoL "limitations", measured as scoring below one standard deviation below the mean on the parent-reported Paediatric Quality of Life Scale at age seven. Le et al.'s findings further demonstrate that these associations persist across childhood, emphasising the impact upon both the child and the demand for healthcare services.

Many mental health difficulties in adulthood have been shown to be persistent from childhood and may decrease social functioning and employment prospects (Belfer, 2008; Green et al., 2004). Law et al. (2009) demonstrate that mental health difficulties of children with SLCN similarly persist into adulthood using the British Cohort Study 1970. Receptive language difficulties are identified as children who score two standard deviations below the mean on the English Picture Vocabulary Test (Brimer and Dunn, 1962) at age five. Mental health is characterised within this study using a non-clinical measure, a composite indicator from four measures at age 34: 1) the Rutter Malaise inventory; 2) a 1-10 indicator of life satisfaction; 3) a binary indicator of feeling control over your life; and 4) a binary indicator of self-efficacy, whether the individual tends to get what they want out of life. Significant associations are found between language difficulties and poorer mental health. The 29-year gap between the language measure and the mental health indication demonstrate an enduring relationship.

Le et al. (2020b) recently conducted a systematic review of the health-related quality of life literature associated with low language. Le et al. emphasise and conclude that the substantial healthcare costs associated with low language include a portion of expenditure on language-focused interventions and a portion of expenditure from additional external healthcare costs, unrelated to the low language. Parental concern was a key predictor of overall healthcare service use (Le et al., 2020b). This could help explain some of the results found in McKean et al. (2017b), where HRQoL was measured on a parent-reported scale. Law et al.'s (2009) findings, however, confirm the presence of lower *self*-reported health for individuals with low language. This suggests a robust association between health and language.

1.3.4.2 The importance of health: Individual and national perspectives

As an individual, health allows all other functions in life, hence its correlation with all other life outcomes (Garcia et al., 2012). A healthy nation, by extension, is able to function better economically as more individuals are capable of working. Furthermore, greater health implies lower healthcare demand, leading to a lower national cost of healthcare.

Health itself is a product of endowments and investments, as with most human capital (Cunha and Heckman, 2006). Health outcomes are vital for estimating long term impacts of early interventions and child development indicators (Garcia et al., 2012; Heckman and Kautz, 2012). Health outcomes are strongly correlated with education, as formalised in the Grossman Model within health economics (Grossman, 1972). This association may be initially rooted in the social inequalities associated with child health (Marmot, 2010), although the social gradient for speech and language is steeper than for other child health indicators.

Health measures are therefore essential for inclusion in any model of life experience. Furthermore, there is significant evidence to hypothesise that the mental health outcomes in later life will be poorer for individuals with SLCN in childhood. Given the associations between SLCN and education and mental health, measures of objective health would further contribute to our understanding of the life experiences of individuals with SLCN in childhood.

Two key contributions of this thesis are identified by the current literature relating SLCN to health:

- Extension of research into later adulthood, beyond age 34, where healthcare demand is highest.
- Extension of research into objective measures of physical health.

1.3.5 Occupational success

Economists have analysed labour market outcomes for different individuals for decades. Wage differentials are modelled using the education and experience of the individual in Mincerian wage equations (Mincer, 1962). Extensions have sought to include broader measures of human capital in order to capture the productive potential of an individual person (Becker, 1964).

For the individual, employment and earnings allow for a higher quality of life through factors such as greater health and consumption of goods. From a broader government perspective, unemployment rates and national gross domestic product are both common indicators of economic success. Furthermore, decreased unemployment and increased earnings would benefit the balance of government budgetary spending. These critical outcomes have been understudied in relation to the development of communication skills.

Employment and earnings have been previously related to a number of factors which are associated with early communication skills. Lower earnings have been noted in adulthood for individuals who: are of low socio-economic status (Solon, 1999); are female (Oaxaca, 1973); have limited cognitive or non-cognitive skills (Heckman and Kautz, 2012); have low education levels (Mincer, 1962); have poor health outcomes in adulthood (van Doorslaer et al., 1997). As earnings are becoming more polarised (Gottschalk and Smeeding, 1997), there is evidence that factors such as academic attainment are become more valuable than ever (Carnevale, 2013).

In addition to these correlated factors, the ability to communicate verbally itself is a valuable skill. Many jobs require communication between colleagues, clients, managers and other stakeholders. The importance of social skills, often including direct measures of communication alongside other non-cognitive skills, has risen substantially over recent decades both for finding employment and succeeding within that role (Deming, 2015; Gallie, 2000).

The following section describes and critiques the literature concerning communication skills and occupational success.

1.3.5.1 Employment

Employers commonly cite communication as a fundamental skill in the recruitment process (Archer and Davison, 2008; Lowden et al., 2011), making the job market even tougher for individuals who struggle to communicate as effectively as competing applicants. According to ‘I Can’ (2017), 88% of unemployed young men display speech and language difficulties, which demonstrates the potential scope of these challenges.

Law et al. (2009) demonstrate, within a large sample longitudinal population study, that individuals with poor receptive vocabulary at age five were twice as likely to experience a period of long-term (>= 12 months) unemployment by age 34 after accounting for a wide range of associated risk factors. This demonstrates that language skills as early as five can predict the later difficulties regarding employment. This finding is supported using samples with a history of developmental language disorders from 16 to their mid-thirties (Whitehouse et al., 2009; Clegg et al., 1999). Conti-Ramsden and Durkin (2012), by comparison, find that employment rates did not differ at 19 years of age, although those who were not employed were less likely to be in education or training as an alternative.

For those who obtain employment, a significant body of evidence suggests differential forms of employment. Adults with a history of language impairment are more likely to take part-time or temporary positions (Whitehouse et al., 2009), to take up vocational employment (Records et al., 1992), or to take up less skilled, lower status roles (Carroll and Dockrell, 2012; Conti-Ramsden and Durkin, 2012; Johnson et al., 2010; Felsenfeld et al., 1994). These differences in employment are likely to suggest lower earnings amongst individuals with either concurrent or prior SLCN, relative to typically developing peers, although this has not been widely researched.

One feature of the literature is that the majority of studies reported are relatively small and clinical in nature (i.e. they focus on those with identified difficulties) (Johnson et al., 1999), often due to legitimate concerns about sample size, time constraints on the study and the individualised nature of effective intervention, although these issues mean that it is often difficult to generalise. The consistency of associations, and the supportive findings within the British Cohort Study 1970 by Law et al. (2009), suggests that these observed group differences are scaled across the population. Given the high prevalence (around 7%, (Norbury et al., 2016)) of DLD, the most common condition researched within these papers, the potential economic costs of these employment outcomes are large.

1.3.5.2 Earnings

Whilst employment has been clearly linked to early speech and language development, as well as concurrent communication skills, very little is known about the earnings potential or related productivity of individuals with speech, language and communication needs who have gained employment.

Intergenerational income persistence – how well a parent's income predicts the income of their child once they grow up – is the primary method economists use to measure social mobility (Major and Machin, 2018). Language skills in early childhood have been shown to capture an important part of the mechanism through which advantage is passed down inter-generationally. Blanden (2007) found that child language at five within low-income families predicted a child's potentially to be upwardly mobile, i.e. to earn more than would be expected based upon their parent's income. Blanden et al. (2013) also demonstrate that the decrease in social mobility over time in the UK is partially explained by an increasing polarisation of both income and childhood skills, including measures of language development. If inequalities grow within one generation in the labour market, this inequality is transmitted into their children's early development of skills and subsequent labour market inequalities. Language skills, amongst other important cognitive and non-cognitive skills in childhood, are an important element in increasing the opportunities available to children from socio-economically disadvantaged backgrounds. Failure to act within the early years may allow inequalities to set, leading to more unequal outcomes in later life.

Blanden (2006) demonstrated that overcoming initial disadvantage, defined as being in a family in poverty, to achieve later life success, defined as their own income in adulthood, was strongly related to cognitive traits measured at age 5 and 10 within the BCS70 including the English Picture Vocabulary Test scores measured at age 5. More recently, Blanden and Machin (2017) used home ownership to model changes in wealth persistence in the UK. They show that a clear association remains between parents and children in terms of their position within the national distribution of wealth. There are multiple transmission mechanisms (see Picketty, 2000 for a full discussion) explaining the maintenance of advantage over multiple generations of a family. Early development of human capital, including cognitive, non-cognitive, and communication skills represents an essential part of the story.

Very few studies have attempted to connect earnings in adulthood to early communication directly. Marsh et al. (2010) conducted a cost-benefit analysis of a speech and language therapy intervention, utilises predicted lifetime earnings as the outcome measure. They estimated that individuals with speech impairment, language impairment or both, at age seven, would earn approximately £34,000 more over the life course as a result of seeing a speech and language therapist, and earn approximately £4,327 more than that as a result of receiving "enhanced" speech and language therapy. They find a 6.43:1 benefit:cost ratio of the enhanced treatment arm. However, this analysis was considerably flawed in many of the extrapolative decisions.

The intervention was estimated to generate a one standard deviation increase in expressive language skills per child, relative to the treatment as usual arm. No confidence intervals or bootstrapping techniques were used to test the sensitivity of this estimate across a population. A total of 152 children were included in the initial intervention study (Boyle et al., 2007) across five treatment arms from a community sample in Scotland. These results were generalised to the broader United Kingdom data for speech and language therapy costs and benefits.

The impact of the intervention on either speech or receptive language was not included in the analysis. A one standard deviation increase in expressive language was then assumed to produce an increased probability of achieving 5 A*-C grades. This does not account for any other possible differences between the children selected for the study at baseline and the children who would have initially scored one standard deviation higher on this measure.

Marsh et al. estimate the lifetime earnings associated with the intervention by estimating the increased lifetime earnings of an individual who achieves 5 A*-C grades relative to someone who doesn't (estimated to be £160,000 from Cummings et al. (2007) with no reported confidence intervals), multiplied by the change in the probability of an SLCN individual achieving this after receiving the intervention. This assumes that the intervention effects are maintained not only past follow-up, at seven years old, onto Key Stage 2 results, but further into Key Stages 3 and 4 and beyond. Limited sensitivity analyses are conducted using one extrapolation at a time in isolation.

Extrapolative models rely upon many assumptions. When taken together, the resulting estimates cannot be interpreted with any great confidence. To the author's knowledge, only two studies have utilised prospectively gathered data to estimate the earnings in adulthood for children who experienced SLCN, one of which was conducted during the composing of this thesis.

Records et al. (1992) report similar income levels between a small sample (n=29) of individuals with DLD and a control group in America. However, these individuals ranged from age 17 to 25, meaning that only a small number of participants were in full-time employment intending to maximise their income, whilst others were in full-time education or part-time employment. This small sample modelled into very early adulthood in the US was not considered a reasonable estimate of the earnings potential of individuals with SLCN for the UK.

Recently, Conti-Ramsden et al. (2018) conducted an analysis of earnings within a UK longitudinal sample, the Manchester Language Study, comparing individuals (n=84) with DLD to a matched control group between ages 23-25. Income was measured categorically in bands. Whilst similar issues apply relating to the age of the participants, they find multiple interesting results.

They report that the DLD group are four times less likely to be in a professional occupation (see Szreta, 1984 for a summary of occupational schema in the UK over the past century). Individuals with DLD report greater difficulty with each of the models of recruitment they had experience of – using a CV, attending an assessment centre, attending a face-to-face interview or having a telephone interview – which helps to explain the prior associations with employment rates.

Income levels were not significantly different between the two groups after accounting for either the type of employment (full-time, part-time, other) or the highest educational qualification obtained in ordered logistic regressions. However, the summary statistics presented in Table 9 of Conti-Ramsden et al.'s (2018) paper show that there is some divergence between the two groups at the higher income levels.

There are four key methodological issues within this analysis which are not addressed by any other papers regarding speech and language development:

Firstly, income is measured only up to age 25. Many economic or sociological studies relating to income assume completion of upwards career trajectories by age 35, but this is not identical across social classes. Specifically, higher paid jobs tend to require more experience, meaning that inequalities may be underestimated within studies of more recent birth cohorts (Gregg et al., 2014; Grawe, 2006). Conti-Ramsden et al. (2018) acknowledge that the participants are in the early phase of their careers and that divergence of incomes may occur later. This thesis utilises alternative data sources to extend our knowledge of earnings into ages where earnings better reflect maximum earnings potential, including measures from age 30 to 62.

Secondly, only one timepoint is used for income. As researchers, we are interested in the lifetime earnings potential of individuals with SLCN and only observing one timepoint leaves the analysis susceptible to temporal biases. This is particularly relevant as many studies have shown that children with SLCN are more prone to periods of long-term unemployment. Comparing two individuals who are not in employment at the time of the questionnaire does not account for how long they have been out of work for, or their probability of gaining employment after the date of the questionnaire. This thesis utilises data with measures of income at multiple ages to minimise the effects of temporary unemployment on our estimates of maximum earnings.

Thirdly, the sample was initially clinically recruited. This explains the relatively small sample size and the resulting difficulty in analysing the small numbers of individuals at the higher income levels where differences may be present but not statistically significant. This recruitment also necessitates a non-normal distribution of scores on the language assessments, with most of the DLD group at the lower end, such that all results are modelled as group mean differences rather than modelling the impact of different facets of language ability as continuous variables.

Finally, the use of categorical bands for income measurement is common in second-hand data sources. However, it limits the financial interpretation of results in understandable terms. Blanden et al. (2013) utilise alternative methods of analysing this data including using the midpoints of the intervals of each category or fitting a Singh-Maddala distribution to the data to simulate an income distribution.

In summary, the available evidence suggests that individuals with early speech and language difficulties do not earn any less than typically developing peers by age 25. However, research extending further into adulthood suggests substantial differences in employability and career paths. Robust analysis of income from at least age 35 is necessary to answer the fundamental question as to whether individuals with SLCN in childhood go on to earn less as adults or not.

The earnings of an individual predict many of their life experiences, including quality of life, stress, propensity to commit crime and use of health care services (Fu and Wolpin, 2013; Mangalore, 2006; Hirsch, 2013; Becker, 1968; Blanden, 2006). Furthermore, the collective earnings of the nation make up the national gross domestic product, reflect the overall productivity of the country and dictate the overall tax receipts used to fund the educational, health care and general services which support the individuals with SLCN to whom this thesis relates.

1.4 Summary

Speech, language and communication needs affect around 7-10% of children at some point in their childhood. Whilst there is instability in the profiles of development within the early years, difficulties still present at school entry are unlikely to later resolve.

Being from a disadvantaged background, however defined, suggests a higher probability of atypical speech and language development. Early attainment deficits for children with communication difficulties at school entry are present at the end of formal education and, although not formally tested, suggest a lower probability of continuation into higher education.

Evidence into adolescence suggests a differential development of non-cognitive personality traits. Children with language disorders demonstrate difficulty in social situations despite highly prosocial intentions. Given the strong associations with both externalised behaviour and internalised mental health problems within individuals with SLCN, further research into the broader positive non-cognitive traits would be beneficial for establishing areas of relative strength and difficulty.

Extrapolated data and prospectively gathered employment data both suggest substantially lower average earnings for individuals with SLCN although this has not been measured using prospectively gathered data.

The lack of evidence extending into adulthood and particularly later life are primarily a result of the lack of easily available data and the focus on clinical populations. This represents a sizeable gap in our understanding of the nature of early communication difficulties.

The following section briefly details the use of the reviewed literature in designing the subsequent four analysis chapters.

1.4.1 The Development of Speech and Language

This literature review summarises the various aspects of communication which may be lacking in an individual with SLCN. Although terminology has changed, many of the descriptions of SLCN within Morley et al. (1954) and Morley (1965) are echoed within Bishop et al. (2017).

This is considered in our interpretation of results in the following analysis, conducted using data collected from 1949-54.

In Chapter 2, Speech, Language and Communication Needs (SLCN) is the term used to describe the group of children identified from these records as having difficulties in any area of their speech and language development. Speech and language are not clearly distinguished due to the historical context of the field at the time (Morley et al., 1954; 1955; Morley, Peckham, 1973). The size of this identified group is discussed in the context of the prevalence estimates within Section 1.2.3. Subgroups of SLCN are identified relating to whether difficulties were 'resolved' by school entry, 'persistent' beyond school entry or 'emergent' after school entry. The relative importance of persistence is discussed in relation to the literature reviewed within Section 1.2.4.

Chapter 3 utilises a continuous measure of 'Receptive Vocabulary' at age five, the English Picture Vocabulary Test (Brimer and Dunn, 1962). Receptive vocabulary has been shown to cross-sectionally correlate strongly with broader language ability (Dionne et al., 2003) and receptive difficulties within this age group are unlikely to later resolve (McKean et al., 2017a; Bornstein et al., 2016). Therefore, this measure is used to represent language difficulties which are unlikely to resolve although the limits of using just one measure are acknowledged.

Chapter 4 continues to use the English Picture Vocabulary Test measure, trichotomizing into a Low, Middle and High Vocabulary group to allow for moderation analysis. Sensitivity of results to changes in the group identification points is tested and reported. This sensitivity is rarely reported despite the clear use of threshold values throughout the literature, as discussed in Section 1.2.3.3.

Chapter 5 also utilises the English Picture Vocabulary Test measure, testing the common threshold values drawn from the literature to test the accuracy of these thresholds as a proxy identification for functional language difficulties. This proxy identification is intended to highlight the poor marginal diagnostic properties of strictly applied thresholds to standardised tests. The importance of thresholds is discussed in more detail within this chapter, in addition to the brief reference to this issue in Section 1.2.3.3. This is not expanded upon within the introduction as Chapters 2 and 3 do not use any threshold values. Chapter 4 does not indicate any diagnoses using threshold values, as many language assessments do (Nitido and Plante, 2020). Chapter 4 does use threshold values to indicate a ranking, as is typical of educational assessments, and reports the sensitivity of the results to changes in the threshold.

Chapter 5 of this thesis critically assesses the validity of strictly applied thresholds for diagnosing functional language difficulties at the margin using a Regression Discontinuity Design. Issues around threshold values are therefore discussed in more detail within Chapter 5 than the short discussion in Section 1.2.3.3.

The risk factors and covariates outlined within this literature review in Section 1.2.5 are considered in each analysis. Chapters 2 and 3 control for available measures of socio-economic status, gender assigned at birth, cognition, birth weight and family size. Chapter 4 directly models the effects of significant risk factors using a decomposition technique (Oaxaca, 1973; Blinder, 1973). Chapter 5 removes the effect of both observable and unobservable covariates using a regression discontinuity design.

1.4.2 The Value of Speech and Language Throughout Life

This literature review emphasises some key gaps in the literature: limited sample sizes, short-term follow ups and a narrow range of methodologies until recently are addressed through the chosen analyses conducted in Chapters 2 to 5. Furthermore, the lack of research into a broader range of outcomes is addressed through the selection of two specific datasets.

Chapter 2 utilises the Newcastle Thousand Families Study (1947) and outcome measures taken in 1997, at age 50, and 2009, at age 62. This dataset contains information on educational achievements from the 11+ examinations until the completion of formal education. Non-cognitive skills are measured using a 60-item inventory parameterising the OCEAN characteristics in order to reliably test the hypotheses outlined in Section 1.3.3 well into adulthood. Health outcomes are measured in both subjective and objective forms for physical and mental well-being, extending the current research in the manner recommended within Section 1.3.4. Finally, income is measured at the household level at age 50 and 62, allowing time for the divergence of incomes to model whether SLCN predicts this divergence.

Chapters 3 to 5 utilise the British Cohort Study (1970) for several major benefits. Firstly, this dataset increases the sample size even further. Secondly, it is a national sample, increasing the generalisability of findings. Thirdly, it is collected 23 years later than the Newcastle Thousand Families Study, increasing the generalisability of the overall thesis discussion.

Chapter 3 focuses on the earnings of individuals with early language difficulties. Earnings are estimated using multiple timepoints. Each of the four issues outlined with the literature within Section 1.3.5.2 are addressed in the design of this analysis. This helps to answer the research question ‘Do individuals with childhood SLCN tend to earn less in adulthood’?

Chapter 4 utilises an econometric method known as the ‘Oaxaca-Blinder Decomposition’ to develop explanations for the earnings gap identified in Chapter 3. Receptive language skills are tested as a moderator of the impacts of the covariates and risk factors discussed in Section 1.2.5 for predicting earnings. Low vocabulary groups exhibit lower average levels of education and cognitive skills, explaining a significant proportion of the observed earnings gap.

Chapter 5 utilises a different econometric method, a ‘regression discontinuity design’. Here, a variety of outcomes are again employed to expand our view of the life experiences of individuals with childhood SLCN. Threshold values are tested for identifying marginal differences in language scores and whether these predict significantly different outcomes at the margin. No differences are observed at specific thresholds, suggesting that strictly applied thresholds cannot identify differences in language ability than functionally impact the life experiences of any individual with SLCN.

Chapter 6 provides a summary and discussion of the thesis, including the literature reviewed, results presented, research designs employed and recommendations arising from the work.

1.4.3 Aims of the Thesis

As reported in Section 1.1.1, this thesis has three fundamental targets addressed by the research design and research questions within each analysis:

1. To utilise historical data to highlight the enduring impact of early of early speech and language development over the life course.
2. To expand the range of econometric methodologies and outcome measures used in the field to increase understanding of the breadth of these impacts.
3. To make appropriate, evidence-based recommendations for research and clinical practice.

Based upon the reviewed literature, this thesis will utilise two cohorts containing information on both early communication development and a wide array of later life outcomes. A range of econometric methodologies are employed to establish the key long-term impacts and the relevant pathways. The discussion makes clear the strengths and limitations of these methods and data whilst making suitable recommendations for researchers and clinicians.

2 Newcastle Thousand Families Study 1947

2.1 Introduction

Chapter 1 notes a lack of later life analyses conducted relating to childhood speech and language development, particularly within population studies due to a paucity of available data. This Chapter parameterises the oldest population-level speech and language development data in existence for the first time since the original participants were re-contacted as adults.

The data contains health visitor records of 887 children born in Newcastle-Upon-Tyne in 1947. These records detail the development of speech and language between the ages of two and seven for each child at regular intervals.

An identifier of early life Speech, Language and Communication Needs (SLCN) is built, including subgroups to capture the resolving, emerging or persistent nature of the child's difficulties.

Accounting for a range of child and family factors observed before adulthood, linear and ordered logistic regressions are used to predict outcomes at age 50 and 62.

A wide range of negative associations are found relating to SLCN, including educational attainment from 11 to adulthood, income at age 50, health at age 50, lower openness to new experiences and lower levels of extraversion. However, negative associations are not universal. Income at age 62, health by age 62, and levels of conscientiousness, agreeableness and emotional stability are not significantly associated with SLCN in childhood. Furthermore, there is significant within-group heterogeneity in outcomes.

The life outcomes for children with SLCN in this cohort are not universally negative, relative to the typically developing group, although there is substantial evidence that this group are less likely to experience educational and occupational success.

2.2 Title

Speech and language development in early childhood and its economic effects into adulthood: A life-course analysis of the Newcastle Thousand Families Study

2.3 Research Questions

1) Are early speech and language development profiles evident in the data?
   a) How does the observed prevalence of childhood Speech, Language and Communication Needs (SLCN) compare to the current literature?
2) What are the effects of SLCN on…
   a) academic trajectories from the 11+ through to the maximum academic attainment observed by adulthood?
   b) earnings in adulthood?
   c) health in adulthood?
   d) the Big 5 personality traits as measured in adulthood?

2.4 Sample

The Newcastle Thousand Families Study (1947), as the name suggests, is a local birth cohort of 1142 of the 1146 children born within Newcastle-Upon-Tyne in May and June of 1947. The study was designed to run for one year to investigate the high infant mortality rate in the city relative to the rest of the UK. The success of this early research inspired successive teams managing the study to maintain contact and trace the entire lives of as many participants as possible.

This seminal study was the first to include data on speech and language in childhood with Morley (1965) representing a landmark publication in the field. This data was collected primarily by health visitors visiting the children's homes approximately once every three months. Annually, Dr Morley would accompany the visit and the data produced forms many of the conclusions about speech and language development reported in Morley (1965).

The sample being drawn from a two-month wide window has the benefit within early child development research of eliminating month of birth effects. This is particularly relevant to the coding of health visitor observations of each study child's speech and language development from the age of 2;0 (2 years, 0 months) to 7;6 (7 years, 6 months) as it means that each child is very close to the same age as each other at formal school entry. This enables a fair comparison of the participants' development of speech and language in their pre-school and early school years. The process of coding this data is described in more detail in the next section (Speech and Language Development Data).

For the first 15 years of life, these children and their families responded to a wide array of questionnaires and assessments annually. The study team were permitted access to the children's health and school records until age 15, representing a very early use of administrative data in research. The range of measures taken cover much of the relevant literature discussed in the prior literature review. Family circumstances, including socio-economic status, early speech and language development, academic achievement including cognitive IQ measures and vocabulary development. These are discussed in more detail in Sections 2.5 Speech and Language Development Data, 2.6 Covariate Measures and 2.7 Outcome Measures.

In 1997, with 35 years having passed since the most recent contact of the full cohort (for a full description of all subsample contact see Pearce et al., 2009), the cohort was followed up at age 50 in 1997. A co-ordinated effort was made to contact the original participants and 89% of the cohort were traced and contacted. Over half of the original cohort (n=574) returned questionnaires detailing their career trajectories and health up to age 50. The majority of these individuals (n=434) also returned questionnaires at age 62 containing information on their current or upcoming retirement experiences, further health and quality of life questionnaires as well as a comprehensive 60-item measure of the "Big 5" OCEAN non-cognitive skills. Clinical examinations were combined with these questionnaires at both adult waves, producing data on objective health for those who attended (sample attrition discussed in more detail in Section 2.8).

Despite the extraordinary length of time over which data has been collected, far pre-dating computerised data management, and the breadth of sources used, the smallest sample for any full information regression still contains 277 individuals. This response rate motivated the recovery of the earliest speech and language records of any cohort dataset to enable analysis of the impacts on these multiple areas of life. These speech and language records are reviewed, parameterised and summarised in the following section.

2.5 Speech and Language Development Data

2.5.1 Observations from age 2;0 to 7;6

The health visitor linked to each study child was asked to observe their development of speech in the early years of life. A visit was made to the child's home approximately once every three months from their 2nd birthday until school entry at age 4;3 (4 years, 3 months). The notes taken by the health visitor were recorded for the study team and subsequently digitised in 2016 for the purposes of this thesis. The health visitor visits became less frequent, occurring approximately once every six months from the onset of schooling at age 4;3 until 7;6.

At ages 3;6, 4;6 and 7;6, a qualified speech therapist linked to the study would accompany the observation to supplement the health visitor's records. In most cases, this speech therapist was Muriel Morley FRCSLT OBE whose work grounds much of the prior literature review (see Section 1.2.2).

The timing of visits was subject to flexibility on the part of both the health visitor and the family and therefore not every child has an equal number of observations at exactly equal ages. For 884 children, there were at least two observations made in both the pre-school years and the school years. This gives a maximum of 24 observations and a minimum of four for each child.

There were three key factors which the health visitors were initially asked to report on at age two:

- Could the child speak?
- Could the child speak intelligibly?
- Was the mother concerned?

These three factors resulted in many reports of the onset development of words and sentences, and subsequently indicating the child's development of vocabulary. In the pre-school years, the reports tended to indicate the onset of spoken language into single words, phrases, and intelligible speech. In school years, there was more focus on intelligibility ("fair", "nasal" and "indistinct" were common phrases), difficulty producing specific consonants (occasionally with specific omissions or substitutions noted, although not always), and vocabulary (not measured on any comparable basis). The most common noted difficulties in the school years were stammers, lisps and dyslalia (phonological speech sound disorder discussed in the literature review Section 1.2.2).

Due to the inconsistency in reporting detail and specific nature of many of these difficulties it was not feasible to generate groups based upon specific impairments. For example, for fourteen children in the identified SLCN group, there are multiple references to them attending a school speech clinic or a speech therapy clinic, but no specific difficulties recorded. It is important to reiterate that the analysis proposed does not seek to establish the later life outcomes of children with specific difficulties in very small group sizes but to provide a more general overview of the outcomes for children with SLCN.

Therefore, the observations were used to represent the presence of <u>any</u> speech, language or communication need experienced by the child considered atypical for their age. This variable is coded as 'SLCN' and is a binary indicator. The following section summarises the separation of development profiles into subgroups, intended to differentiate persistent SLCN from resolving SLCN (see Section 1.2.4 Stability of Speech and Language Development).

2.5.2 Age of Difficulties

The difficulties observed were separated by age using the school entry age of 4;3 as a cut-off. This is in reference to the literature (see Section 1.2.4) on the stability and effects of early communication difficulties which suggest that those disorders not resolved by school entry are unlikely to resolve without targeted intervention and are likely to have larger impacts on long term outcomes (see Durkin et al., 2009 for an example using resolving vs persistent specific language impairment to predict GCSE outcomes relative to a typically developing control group, also discussed in Section 1.3.2 Education and the Attainment Gap).

The health visitor data contained a Yes/No category for speech and language difficulties reported on the child's school card. Children reported to require intervention using this measure were categorised as having observed difficulties in the early primary school years (4;3 to 7;6) as this report card was not dated but was contained within the health visitors records. However, the school report card did not identify any SLCN cases which were not also noted by the health visitor.

2.5.3 Subgroups of SLCN identified

Four subgroups are created within the overall sample (rounding to one decimal place, hence the 99.9% total):

1) *TYPICAL* – No Difficulties / No Difficulties: Children with no observed difficulties or minor age-appropriate difficulties at any age were coded as 'Typical' (n=725, prevalence = 81.7%.[1]
2) *RESOLVING* – Difficulties / No Difficulties: Children with observed difficulties in pre-school years which resolved by school entry were coded as 'Resolving' (n=62, prevalence = 7.0%).
3) *EMERGENT* – No Difficulties / Difficulties: Children with no noted difficulties in pre-school years but with difficulties noted in the school years were coded as 'Emergent' (n=19, prevalence = 2.1%).
4) *PERSISTENT* – Difficulties / Difficulties: Children with difficulties observed in both the pre-school and school years were coded as 'Persistent' (n=81, prevalence = 9.1%).

Overall, children with any considerable SLCN (n=161) notes from age 2;0 to 7;6 represent 18.2% of the sample. Half of the SLCN sample displayed difficulties in both pre-school and school early years. Only a small minority of difficulties emerged after school entry.

Figure 2.1 demonstrates the trajectories of each group from early years (2;0 to 4;3) to the early school years (4;3 to 7;6). Table 2.1 summarises the above subgroupings and sample sizes.

The majority of pre-school difficulties resolve by school entry with a large minority remaining persistent. A small minority of those without any difficulties noted before school entry develop later difficulties noted by the health visitor or schoolteacher.

Two example cases of each subgroup are presented below in Tables 2.2a-h.

[1] Rounding to one decimal place leads to the total slightly under 100%.

Table 2-1 Subgroups of speech and language development profiles and sample sizes

Subgroup Profile	Pre-School (2;0 – 4;3)	Early School (4;3 – 7;6)	N
Typical	Typical	Typical	725
Resolving	Difficulties	Typical	62
Emergent	Typical	Difficulties	19
Persistent	Difficulties	Difficulties	81
Total			**887**

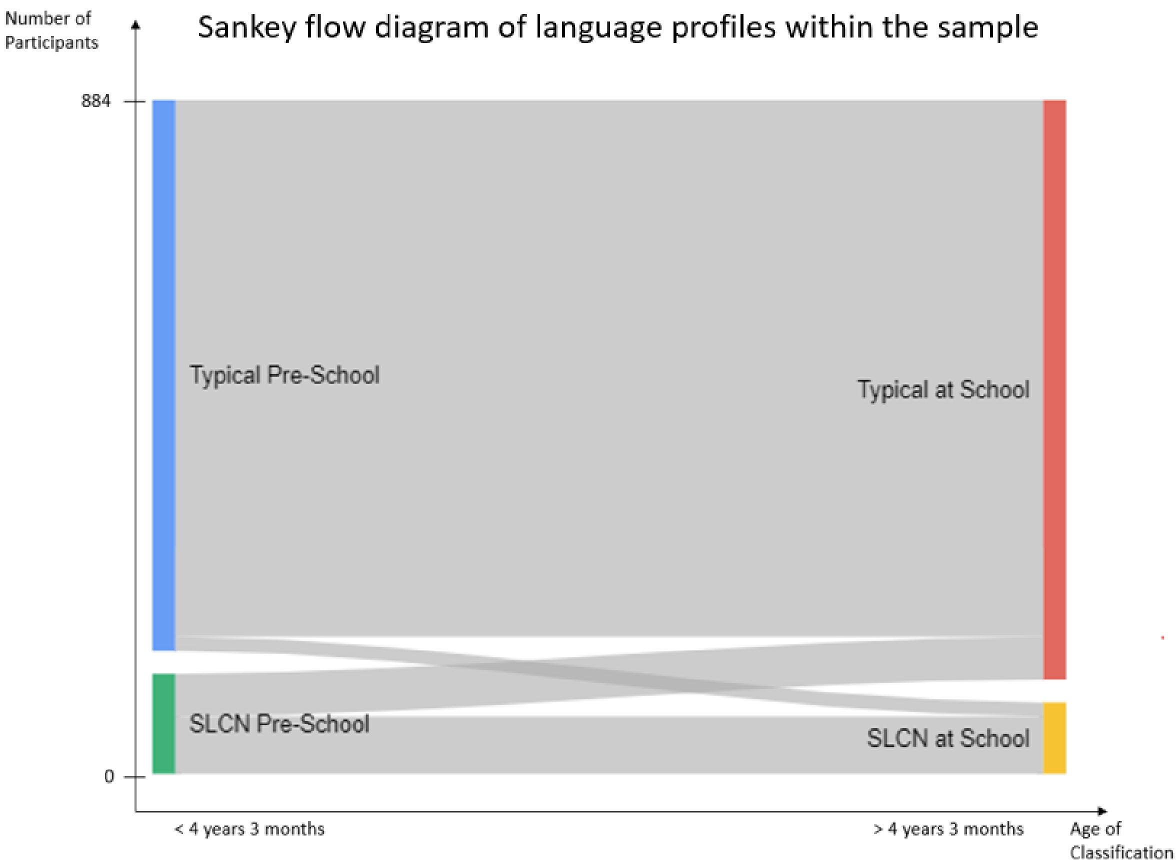


Figure 2-1 Flow Diagram of Sample into Speech and Language Development Subgroups

Note: Flow diagram consists of the four subgroups, 'Typical' (Blue to Red), 'Resolving' (Green to Red), 'Emergent' (Blue to Yellow) and 'Persistent' (Green to Yellow).

2.5.3.1 Examples of Speech and Language Development Profiles

The date of observation is noted above each observation, dated chronologically left to right, top to bottom. The first observation takes place near child's second birthday. The red observation signifies the first observation following school entry.

Table 2-2a Typical speech and language profile example one

Date	03/06/1949	19/08/1949	16/11/1949	19/01/1950	07/03/1950	04/05/1950
	Good	Very Advanced	Satisfactory	Very clear	Satisfactory	Satisfactory
Date	27/07/1950	27/09/1950	06/12/1950	04/04/1951	23/05/1951	31/06/51
	Satisfactory	Satisfactory	Satisfactory	Satisfactory	Satisfactory	Satisfactory
Date	01/08/1951	30/10/1951	20/12/1951	27/02/1952	06/05/1952	31/07/1952
	Satisfactory	Satisfactory	Satisfactory	Satisfactory	Satisfactory	Satisfactory
7th yr	22/07/1953	20/10/1953	17/12/1953	04/03/1954	10/06/1954	
	Satisfactory	Satisfactory	Satisfactory	Satisfactory	Satisfactory	

No notes of concern made across 23 observations, 13 before school entry and 10 more following school entry.

Table 2-2b Typical speech and language profile example two

Date	31/08/1949	10/11/1949	24/01/1950	24/04/1950	13/09/1950	12/12/1950	09/03/1951
	Good	Good	Good	Good	Good	Good. Occassional difficulty with initial 's'	Satisfactory
Date	26/09/1951	18/02/1952	21/05/1952	31/10/1952	30/01/1953	29/09/1953	
	Satisfactory	Satisfactory	Good	Normal	Normal	Normal	
Date	10/02/1954	21/09/1954					
	Normal	Normal					

One age-appropriate difficulty mentioned with no other notes of concern made in 14 other observations.

Table 2-2c Resolving speech and language profile example one

Date	05/07/1949	29/08/1949	03/11/1949	03/01/1950	09/03/1950	15/05/1950
	1 or 2 words - tea, ma ma etc. Will try to imitate but the sounds are very indistinct. Grunts when she wants to attract attention	Few more words eg apple, averalls (very indistinct)	Can say several odd words now but has not attempted to string any together yet	The same	Odd words	Improving- small sentences
Date	27/07/1950	23/10/1950	28/03/1951	25/05/1951	25/05/1951	08/10/1951
	Large vocabulary	Good	Good	Good	Good. Babyish but clear	Good
Date	13/12/1951	11/03/1952	30/06/1952	23/09/1953	28/07/1954	
	Good	Good. Intelligent	Good	Normal	Normal	

Early difficulties with pronouncing words clearly and delayed onset of connected speech. Small sentences appear at age 3;0 and from there the majority of observations suggest positive development with no concerns noted after school entry.

Table 2-2d Resolving speech and language profile example two

Date	16/06/1949	09/07/1949	08/09/1949	16/11/1949	08/02/1950	24/03/1950
	Fairly good	2-3 words together - not advanced	A bit slow	Fair	Indistinct	Fairly Good
Date	30/05/1950	24/08/1950	04/01/1951	01/11/1951	23/01/1952	30/10/1952
	Fair	Fairly good	Speech therapist visited last week but ____ did not speak	Good	Normal	Satisfactory

Slow development of piecing words together and indistinct speech. Speech therapy noted aged 3;8 with the child not speaking. All observations after school entry suggest resolution of these difficulties although data is limited with only three observations recorded after school entry.

Table 2-2e Emergent speech and language profile example one

Date	15/09/1949	14/11/1949	26/01/1950	22/03/1950	19/06/1951
	Good	Good	Good	Good	Normal
Date	18/06/1951	12/09/1951	06/02/1952	16/09/1952	20/05/1953
	Normal	Normal	___ is a poor speaker, will not speak to Health Visitor but was heard speaking to other children	Improved, much less shy	Normal
Date	28/03/1953	09/10/1953	08/12/1953	14/01/1954	02/07/1954
	Still difficult	Improved	Difficulty with 'm', nasal speech	Still has difficulty has been seen at school clinic	Still some difficulty

Seven observations made with no concerns noted from age 2;3 to school entry at 4;4. Before fifth birthday, they were described as a "poor speaker", refused to speak to the health visitor and specific difficulties noted from age 6;6 including attending a school clinic for speech development.

Table 2-2f Emergent speech and language profile example two

Date	12/09/1949	30/11/1950	19/03/1951	22/08/1951	19/03/1952
	Satisfactory	Good	Bright	Good	Speaks as if he has a nasal obstruction
Date	27/08/1952	29/04/1953	10/11/1953	04/02/1954	15/07/1954
	Fairly good. Some words not distinct	Lisps	Lisps	Not clear	Some words not clear. Dyslalia

Four observations made at significant intervals with no negative notes made on the child's speech and language development prior to school entry. Following school entry, specific notes are made on a variety of difficulties. 'Dyslalia', as shown in Table 1.1, is a specific diagnosis from this time period relating to phonological development where articulation development is outstripped by vocabulary development, generally not diagnosed until after school entry.

Table 2-2g Persistent speech and language profile example one

Date	10/11/1949	17/01/1950	16/03/1950	18/05/1950	06/07/1950
	Not very clear. Plays with brother who is not very distinct and probably imitates	a little improved	Indistinct. Copies brother and doesn't finish word	improving	Backward
Date	06/09/1950	14/11/1950	12/03/1951	26/07/1951	15/08/1951
	no apparent improvement attending speech clinic	still backward	slight improvement - to attend speech clinic after easter	as before	mother does not see improvement but friends do
Date	17/10/1951	30/11/1951	29/02/1952	N/A	29/05/1952
	very difficult to understand - still visiting speech clinic	unchanged	mother does not see improvement but friends do	Attends speech therapy	still defective. Mother says better some days than others
Date	24/10/1952	30/10/1952	13/08/1953	29/09/1953	06/07/1954
	greatly improved	improving but still consonant defect. Teacher says he speaks perfectly and is in top group for reading	improving	appeared normal unless he is excited	satisfactory

In this example, the difficulties are noted from the first observation and the speech clinic is attended from at least age 3;4. There is noted improvement after school entry although consonant defects are still noted at 5;5 after over a year at school. The terminology used is reflective of the era and does not represent modern diagnoses of speech, language and communication needs.

Table 2-2h Persistent speech and language profile example two

Date	01/06/1949	28/02/1951	09/05/1951	13/07/1951	21/09/1951	29/11/1951
	Quite well. 2 Words together.	Stammers, can't pronounce 's' or 'f'	Stammers	Stammers a little. Improving	Fairly satisfactory	Lisps and has light stammer.
Date	04/02/1952	24/04/1952	11/07/1952	28/10/1952	24/12/1952	17/04/1953
	Very fair	Much better. Stammers very little now.	Defective	Defective	Defective	Much better. Does not stammer now
Date	21/05/1953	09/07/1953	02/10/1953	10/12/1953	11/02/1954	13/05/1954
	Not good. Needs therapy but mother has not time to take him	Very Poor	Very defective. Numerous appointments with SLT none of them kept.	Defective	Can't pronounce some consonants.	Satisfactory
Date	31/08/1954					
	Dyslalia					

Although the child developed language within a normal developmental timeframe, a stammer was noted at 3;9 alongside two consonant difficulties. This stammer has ceased by age 5;11 although consonant defects persist and a diagnosis of dyslalia is given at age 7;3.

2.5.3.2 Relation to common risk factors

The 887 study members for whom speech and language data was collected and analysed were evenly split by gender (444 males and 443 females). However, 63.6% of the SLCN group were male, a statistically significant difference. This suggests male: female prevalence ratio of 1.75:1 which is consistent with those observed in the literature which range from 1:1 (Tomblin et al., 1997) to 2:1 (Leonard, 2014). The presence of expressive speech difficulties, as opposed to a specific focus on language, within our sample may explain the gender ratio (Wren et al., 2016) which is relatively large for a population study (see Section 1.2.5.2 for more detail on this).

To make a simple comparison of SES groups, we dichotomise the father's occupational social class into Low SES (Unskilled or semi-skilled occupations) and High SES (Technical, managerial and professional occupations). Children from the low SES group (prevalence of 23.3%) were 2.3 times as likely to be in the SLCN group as children from the high SES group (prevalence of 10.0%). This suggests that speech and language development was highly sensitive to social disadvantage within this cohort. Locke et al. (2002) and Law et al. (2017) report that low SES approximately doubles an individual's risk of atypical language development. The slightly higher estimate observed within the Newcastle Thousand Families Study may reflect high levels of inequality than the previous literature, possibly reflecting within-region disparity in this period. However, this estimate could just as easily reflect health visitor reporting bias. These two conflicting hypotheses cannot be formally tested. Furthermore, the majority of participants' father's occupational class was classified as skilled manual (reflective of the labour market in Newcastle at the time), rather than into the high or low SES groups, which limits the sample size available to make comparisons between different occupational groups.

2.5.4 Frequency of health visitor observations specifying difficulties

Even within subgroups, there can be significant heterogeneity in the extent and persistence of observed difficulties. Furthermore, there was no standardised approach to the observation recordings meaning that the level of detail is highly variable. Measuring severity is therefore subjective to both the current thesis author and to the original recordings made by the health visitor or speech therapist.

Therefore, the frequency of difficulties was considered to be the best objective proxy available for severity, strongly related to persistence. This was measured simply as the number of observations made with concerns about the specific child noted, referred to in the results tables as “Frequency”.

For the majority of children in the sample, this was zero. For a child with any non-Typical speech and language development profile, at least one observation must have contained information on some difficulty relating to the child.

Table 2.3 and Figure 2.2 summarises the distribution of the “Frequency” of concerns noted, ranging from zero to 24, separated by speech and language development profile. Clearly, the ‘Persistent’ profiles contain the highest number of negative observations on average due to the length of time over which their difficulties occurred.

For each of the examples noted in Tables 2.2a-2.2h, the number of observations containing noted difficulties is summarised in Table 2.4.

Table 2-3 Frequency summary statistics

Subgroup Profile	n	Mean Frequency	Std. Dev.	Min	Max
Typical	725	0.40	1.18	0	9
Resolving	62	4.19	3.53	1	16
Emergent	19	3.42	2.52	1	9
Persistent	81	10.09	5.72	1	24
Total	**887**	**1.62**	**3.66**	**0**	**24**

Note: ‘Frequency’ is the number of observations by health visitor or speech therapist containing noted concerns on the child’s difficulties with development of speech and language.

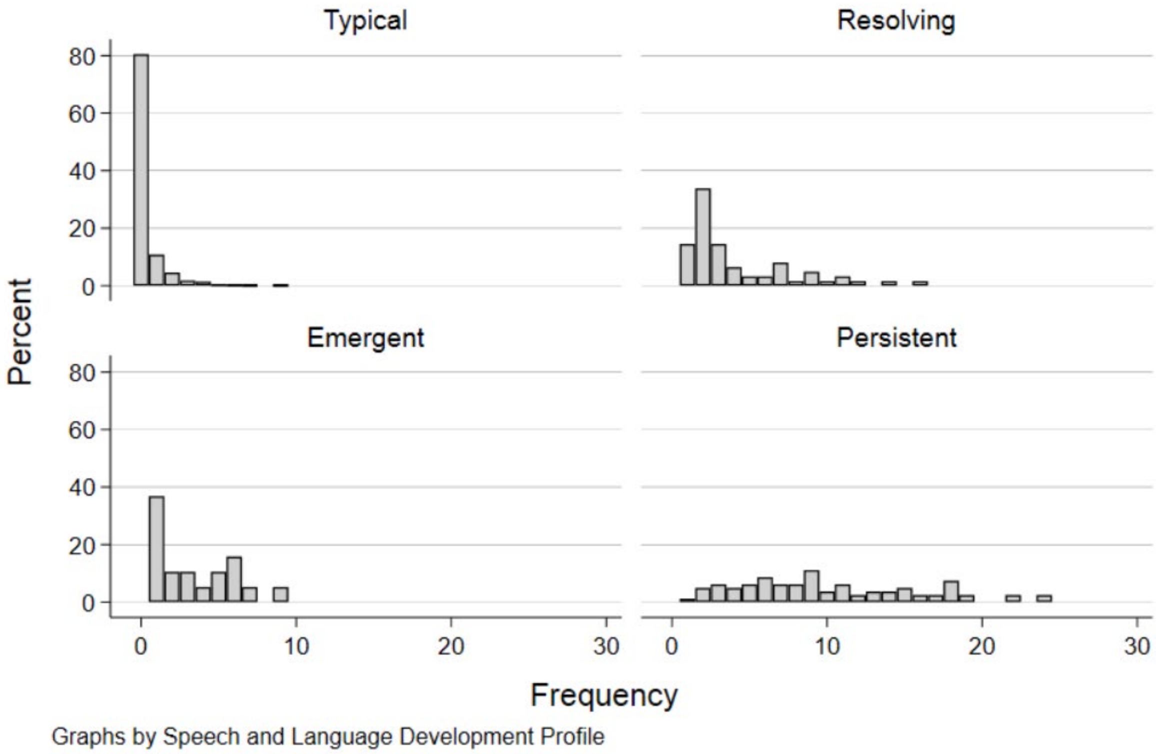


Figure 2-2 Distribution of 'Frequency' variable by SLCN subgroup profile

Table 2-4 Frequency calculations within example speech and language development profiles

Subgroup Profile	Example Number	Table	Frequency
Typical	One	2.2a	0
Typical	Two	2.2b	1
Resolving	One	2.2c	7
Resolving	Two	2.2d	6
Emergent	One	2.2e	5
Emergent	Two	2.2f	6
Persistent	One	2.2g	13
Persistent	Two	2.2h	14

'Frequency' is the number of observations by health visitor or speech therapist containing noted concerns on the child's difficulties with development of speech and language from 2;0 to 7;6.

2.5.5 Research Question 1

1) Are early speech and language development profiles evident in the data?
    a. How does the observed prevalence of childhood Speech, Language and Communication Needs (SLCN) compare to the current literature?

887 study members are classified according a speech and language development profile. 162 of these are identified as having some atypical development in either pre-school years or in the first years of schooling. This is a prevalence of SLCN at any age from 2;0 to 7;6 of 18.2%. As this can emerge or resolve at any time within this range, we expect a higher prevalence rate than those reported from a cross-sectional examination of an age group at one timepoint.

Law et al. (2017) report a median estimate of 14% within two-year olds and a 10.7% median estimate within four to five-year olds within population analyses. Higher estimates would be expected within socio-economically disadvantaged areas. These prevalence estimates are discussed in greater detail in the literature review.

Estimates of SLCN in pre-school children, aged 5;0 and under, range from 2% (lowest estimate found in Law et al., 2000) to 20.7% (Reilly et al., 2010). The prevalence rate in this study of pre-school difficulties is 16.1% which is at the higher end of this range of established estimates.

Within the 18.2% of participants who are identified with SLCN, three distinct subgroups are identified as 'Resolving', 'Emergent' or 'Persistent' difficulties according to the timing of their problems, i.e., in the pre-school years (<4;3) or the early years of school (>4;3). The reported prevalences are 7.0%, 2.1% and 9.1% respectively.

Table 2-5 Prevalence of SLCN within the sample relative to a comparable study

Current Analysis	**'Typical'**	**'Resolving' SLCN**	**'Emergent' SLCN**	**'Persistent' SLCN**
Prevalence Estimate (%)	82.8	7.0	2.1	9.1

Law et al. 2012	***Typical Language***	***Resilient Language***	***Increasingly Vulnerable Language***	***Consistently Low Language***
Prevalence Estimate (%)	*92.7*	*4.4*	*1.4*	*1.5*

Note: Prevalence of four SLCN subgroups compared to similar subgroups of language impairment from three to five years of age in Law et al. (2012).

Very few studies have estimated a sample at multiple timepoints to analyse emergence and persistence (Law et al., 2017a). One example is Law et al. (2012) from the Millennium Cohort Study, a population study, with change modelled from age three to five using a 1.5SD threshold on the naming vocabulary test of the British Ability Scales (Elliott et al., 1997). They report four groups, comparable to our subgroups. Table 2.5 above compares the prevalence rates within our trajectory profiles of speech and language development to the equivalent profiles within Law et al. (2012).

The comparison study is reported as it is a rare example of comparable subgroups within this age range. The estimated prevalence in the comparison study is notably low relative to the literature. Table 3.1 in the Early Education Foundation report (Law et al., 2017a) of early language development shows that the Law et al. (2012) method reported the lowest prevalence at both age three and at age five of any study evaluated. Therefore, a higher prevalence in the current data may not reflect an oversensitive inclusion criterion.

Therefore, each individual subgroup also shows a lower prevalence. Notably, the 'Emergent' category is smaller in my study relative to other subgroups in Law et al. (2012) which is likely to reflect the later age of emergent difficulties (>4;3 in my study as opposed to >3;0). In a study using a threshold cut-off (-1.5SD in Law et al., 2012a), you are likely to have emergence and resolution of difficulties categorised for relatively stable children making only narrow changes around the threshold, which may be due to measurement error (McKean et al., 2017b).

The subgroups identified suggest a relatively high proportion of persistent SLCN difficulties relative to resolving SLCN difficulties. In the identified subgroups within this study, 43% of those with initial difficulties in the pre-school years showed no difficulties in the school years and were categorised as ‘Resolving’. 57% of those with initial difficulties had ‘Persistent’ SLCN into school years. Law et al. (2017) summarise that approximately 70% of impairments identified in pre-school are resolved by school entry whereas only 30% would be expected to persist from the available literature.

It is not possible to test specific hypotheses as to why this is the case. Many possible factors may contribute; health visitor reports are not a common indicator in the literature although differences in prevalence from parental reports and from child performance have been previously noted (Law et al., 2017a); the associated health visitor may be more sensitive to relatively minor later difficulties within children for whom they had previously noted SLCN difficulties in pre-school; without the use of a standardised test, the normal range of performance is subjective and much larger changes may be necessary for the health visitor to state that the SLCN has resolved; the majority of the literature focuses on language difficulties rather than broader SLCN which may be more likely to resolve within early years; intervention may be more effective in targeting and resolving early difficulties than in the 1950s when speech therapy (language not yet specified) was a relatively new and evolving field of healthcare; children born in May or June, as in the Newcastle Thousand Families Study, are relatively young for their school year and therefore less likely to resolve by school entry than children who may be nine months older developmentally.

Finally, this study identifies a measure of persistence and severity proxied using the frequency of health visitor concerns, i.e., the number of observations containing noted difficulties relating to the child. There is not a comparable measure in the literature as this type of second-hand data pre-dating current standardised speech and language assessments has not been previously attempted. Furthermore, variability in the number of total observations limits the inference we can make from this measure.

Prospectively gathered longitudinal individual-level data on the communication development of a large sample of children is vital to improving our understanding of how speech and language develops (Desmarais et al., 2008) and these three variables represent a contribution to the literature on prevalence of SLCN.

2.5.6 Summary

The speech and language development of 887 children was coded into three variables, each of which is used in the later analysis.

- A binary indicator of whether the child's speech and language development were deemed 'Typical' or 'SLCN' (n = 725 and n = 162).
- A subgroup indicator of three types of atypical speech and language development (SLCN), the 'Resolving', 'Emergent' and 'Persistent' subgroups (n = 62, n = 19 and n = 81).
- An indicator for the 'Frequency' of difficulties, the number of observations containing concerns made by the health visitor (range from zero to 24).

Each of these variables is modelled in our analysis to allow for potential correlations with SLCN to vary by timing, persistence and severity.

We can identify speech and language development profiles within the data although there are limitations to its comparability with the literature. The nature of the data is unique and formalising these measures requires simplifications from the complex and varied original recordings.

Estimates associated with these categories would not be extrapolated cleanly onto an alternative measure in other cohorts which are commonly standardised tests taken at one or multiple time points measuring one specified aspect of communication. The measures of SLCN presented, particularly as a binary indicator, are highly inclusive relative to the literature as it focuses on broader communication difficulties than the recent focus on DLD specifically and contains a larger percentage of the overall sample than a strict threshold cut-off would contain. Inclusivity is recommended in Law et al. (2017) as differences in outcomes have been observed in the past using the lowest 10% or 16% of scores on a language assessment as the identifying indicator (Tomblin, 2008; Beitchman et al., 2001).

Prospectively gathered longitudinal individual-level data on the communication development of a large sample of children is vital to improving our understanding of how speech and language develops (Desmarais et al., 2008) and these three variables represent a contribution to the literature on prevalence of SLCN. More importantly, this identification allows for a mapping of the outcomes of a group of SLCN children relative to typically developing children within the longest-running cohort data set to contain any measure of the development of early speech and language.

The longevity and detail contained within the Newcastle Thousand Families Study (1947), as outlined and discussed in the following 'Measures' section, is a relative strength, extending into later adulthood unlike most longitudinal studies. Whilst the measure of speech and language development does not create easy comparability with other measures in the literature, the level of detail over multiple formative years represents a significant strength. The following section summarises the sample attrition over this significant period and number of variables. This is useful for summarising the subsequent sections covering the measures included for analysis.

2.6 Attrition

The sample diminishes in full-information size from its inception in 1947 to its final wave in 2009. This attrition over time is detailed in Figure 2.3 for participant drop-out and is unlikely to be at random.

The initial records of this cohort (Spence et al., 1954) report the increased likelihood of children from disadvantaged backgrounds dying within the first year of life, primarily due to acute infections which were common in the city at the time. There is also attrition from the most advantaged backgrounds because of increased probability of leaving Newcastle in the early years. In the follow up waves in adulthood, residence in Newcastle Upon Tyne was not a necessary criterion.

The sample was recruited from the borders of Newcastle-Upon-Tyne as of 1947 which did not include various affluent parts of the city which were later integrated. For this reason, the conclusions of the analyses presented within this data cannot be considered representative of any specific population. However, the sample does contain a wide range of individual and family circumstances.

In the later cohorts, the sample response rates are slightly higher amongst females and participants with higher 11+ scores. Given that the initial population was not representative of a particular current region, given the border changes in the intermittent 73 years since the study’s inception, survey weights are not applied to reweight the remaining sample to resemble the initial sample.

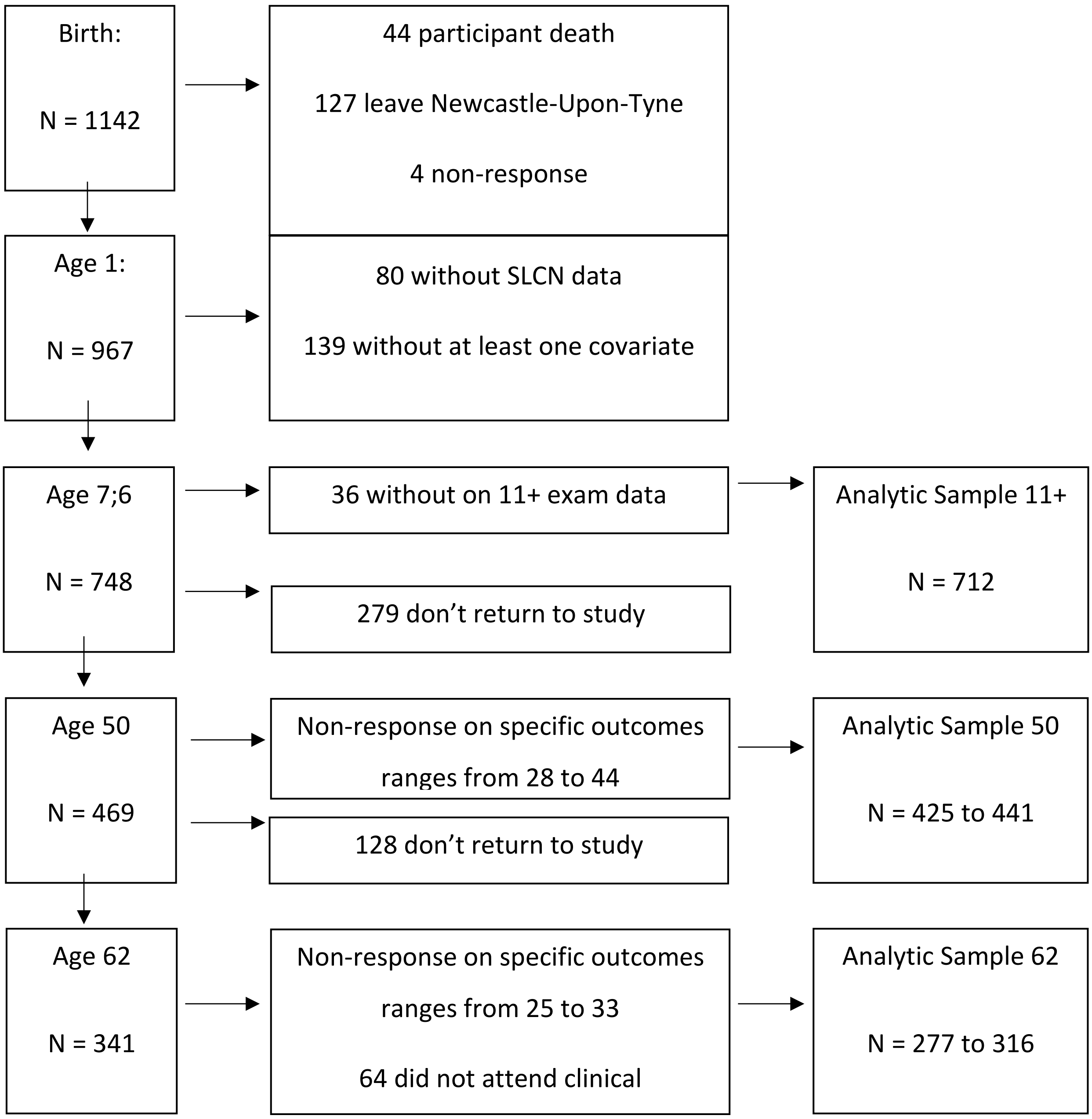


Figure 2-3. Flow Chart of Attrition from Newcastle Thousand Families Study

Table 2-6 Attrition rates over analysis samples between Typical and SLCN groups

	Sample			
	Age 7;6	**Age 11**	**Age 50**	**Age 62**
Typical	725 (81.7%)	579 (81.3%)	383 (81.7%)	284 (83.3%)
SLCN	162 (18.3%)	133 (18.7%)	86 (18.3)	57 (16.7%)
Total	**887 (100%)**	**712 (100%)**	**469 (100%)**	**341 (100%)**

Note: Number of respondents reported within each sample (see Figure 2.3 on attrition and analytic samples) reported with the percentage of the sample identified with SLCN.

There is no significant relationship between SLCN identification or subgroup and attrition with the proportion of the sample identified remaining constant across our analytic samples. This is summarised in Table 2.6 for the SLCN group.

The following section summarises the covariates and outcomes included in the analytic model, making specific reference to the literature reviewed in Chapter 1.

2.7 Covariate Measures

The list of measured covariates is summarised in Table 2.7 with descriptive statistics of each.

2.7.1 Gender

The cohort member's gender assigned at birth was recorded as a binary variable. For our analysis, female was treated as a dummy variable in multi-variate models. Gender is commonly associated with the prevalence and presentation of SLCN in both this cohort and the literature. Within this cohort, gender is correlated with education, health, occupational and personality outcomes as in the literature. Subsequent restricted sampling approaches were taken to model differences in the observed correlations for males than females. Our initial sample was 51% male and 49% female. However, after later attrition to the final analytic sample, the sample became weighted slightly towards females (55%) as discussed in the previous section.

2.7.2 Father's Occupational Social Class

Father's Occupation is recorded at two time points in childhood: at birth and at age five. These were measured using the Registrar General 5 Class system in place at the time. These categories are unskilled, partly skilled, skilled (manual and non-manual), technical and professional. At age five, the skilled category was further separated into non-manual and manual using the 6 Class system (the same as is used in the British Cohort Study 1970 in later analyses). Social class at birth was used to maximise sample size. There were 20 cases where social class was missing at birth and present at age five. We assume that class status is time invariant over this five-year period, as is the case for the majority of participants who respond at both time points and use the identified social class at age five for these cases.

The most common occupational class was skilled labour, particularly manual, reflecting the high proportion of ship builders and similar trades dominating the local labour market in Newcastle at the time (Spence et al., 1954; Miller et al., 1960). The study team regularly contacted the parents within the first year of life and none of the fathers were unemployed for the whole year. The study team recorded the most common occupation of the father across the first year of their child's life. Only 7% of mothers in the sample worked outside the home, hence the focus on father's jobs to represent occupational social class. Unfortunately, no measure of income or parental education level was taken to provide a more thorough representation of socio-economic status for the family.

2.7.3 Siblings

The number of children older than the cohort member within the family was recorded at birth. The number of older siblings was a potentially important covariate due to the relationship with early language development discussed in the literature review. The largest family contained 13 children including the cohort member. The average family within the sample contained one older and one younger sibling. A quarter (n=293) of the sample were only children. The number of older siblings was found to be related to identification of SLCN more strongly than the total number of siblings or a binary indicator of being a singleton within this cohort. All findings are robust to the alteration of this variable to categorical and binary structures.

2.7.4 Childhood Adversity

A scale of childhood adversity was recorded containing six possible adverse events in the child's life before age 15. These are comparable to the modern literature on ACE (Adverse Childhood Events) and are a sum of the six binary measures of adversity. The six events are as follows; household/family in debt or had financial problems, parents were separated, a parent with chronic sickness/illness, a parent involved with the police, the NSPCC were involved with the family and/or the death of a parent. These six binary variables added to create a possible 0-6 scale, although five was the highest score within the sample. 60% of the sample experienced none of the measured events indicating childhood adversity and a further 28% experienced just one. This variable is modelled as a continuous variable, preferred to a categorical variable due to the low sample size in higher value groups.

2.7.5 Birth weight

The child's birthweight was measured and standardised (mean = 0, standard deviation = 1) according to gender and gestation period. Low birthweight has been found to predict early school readiness (Camacho et al., 2019) and cognitive function by age 11 (Drillien, 1967) although evidence from the Newcastle Thousand Families Study has suggested that birth weight is not related to childhood IQ at age 11 (Pearce et al., 2005).

2.7.6 Summary

The list of measured covariates is summarised in Table 2.7 with descriptive statistics of each. The correlations of each covariate with each measure of SLCN are presented in Table 2.8, demonstrating the need for each covariate to be included in our analysis.

Table 2-7 Summary statistics of covariates by analytic sample

Covariate	Analytic Sample	n	Mean	Std. Dev.	Min	Max
Female Dummy	11	712	0.500	0.500	0	1
	50	469	0.551	0.498	0	1
	62	337	0.549	0.498	0	1
Older Siblings	11	712	1.438	1.886	0	11
	50	469	1.447	1.902	0	10
	62	341	1.279	1.784	0	10
Standardised Birthweight	11	712	0.027	0.981	-3.362	4.482
	50	469	0.051	0.988	-3.362	4.482
	62	341	-0.101	0.967	-3.049	2.972

	Father's Occupational Social Class			**Adverse Childhood Events**			
	Analytic Sample				**Analytic Sample**		
	11	**50**	**62**		**11**	**50**	**62**
Professional	10	7	8	**0**	429	270	207
Managerial/Technical	42	30	26	**1**	206	148	104
Skilled (Manual or Non-Manual)	413	293	216	**2**	61	40	23
				3	14	8	5
Partly Skilled	120	69	48	**4**	1	2	2
Unskilled	127	70	43	**5**	1	1	0
Total n	**712**	**469**	**341**	**Total**	**712**	**469**	**341**

Note: Categorical variable distributions are displayed in the latter half of the table vertically.

Table 2-8 Pearson's correlation matrix of all covariates against all three SLCN measures

	Gender	Social Class	Siblings	Childhood Adversity	Birthweight
SLCN (Any)	0.128***	0.105***	0.090***	0.126***	0.079**
Resolving	0.070**	0.001	0.013	0.065*	0.022
Emergent	0.008	0.061*	0.032	0.016	0.012
Persistent	0.105***	0.109***	0.116***	0.103***	0.081**
Frequency	0.099***	0.130***	0.130***	0.158***	-0.014

*Note: Values show linear correlation r (range -1 to 1). * = p < 0.1, ** = p < 0.05, *** = p < 0.01.*

As shown in Table 2.8, these covariates significantly predict at least one of the measures of SLCN used in our analysis as the literature suggested. The Persistent SLCN group show the strongest associations with the covariates of the subgroups and they may therefore control for more of the estimated effect of Persistent SLCN on each outcome.

As shown in the attrition section, none of the covariates have high levels of missing data. Each regression, one per combination of outcome and SLCN measure, will include every covariate. Only the adjusted multivariate estimates are presented in our results tables.

2.8 Outcome Measures

A range of outcomes are modelled to quantify the range of long-term differences and similarities between groups with and without typical development of speech and language.

These fall into four main categories, as formalised in the research questions (2a-2d), as follows:

- Education
- Occupational
- Health
- Personality

A consistent analytic strategy is applied to each of these outcomes, meaning that the same covariates are included in each analysis, and all measures of SLCN are analysed.

2.8.1 Educational Measures

In the primary analysis presented in the Results section, there are three outcomes associated with academic achievement.

2.8.1.1 11+ Score

Each study member's performance on the 11+ exam was recorded as the sum of three sub-scales:

- Verbal IQ (Moray House 57 and 58 measuring verbal reasoning)
- English
- Arithmetic

Each of these sub-scales were combined to give an overall 11+ score which is standardised (mean = 0, s.d. = 1) for the analysis.

The individual sub-scales were tested separately (each standardised) as a supplementary check (discussed in the Results section) using the same analytic approach as the overall score.

The students with the highest scores across the three scales tended to be selected to attend grammar schools and the students with the lowest scores tended to attend secondary moderns. There was a third category of 'Technical' schools with a more vocational focus and these students often showed more variable performance across the three scales, suggesting specific strengths and weaknesses that these schools may have targeted (Miller et al., 1974). This negative correlation between the three sub-scales within technical school students has been noted in previous research (Pearce et al., 2009) and hypothesised to reflect a potential fail grade on one specific scale denying entry to the local grammar school.

There was no specific cut-off identifiable for attendance of a specific school type due to differences in local availability of schools dictating the acceptance of students. The range of average scores across the three scales are presented in Figure 2.4, separated by school type.

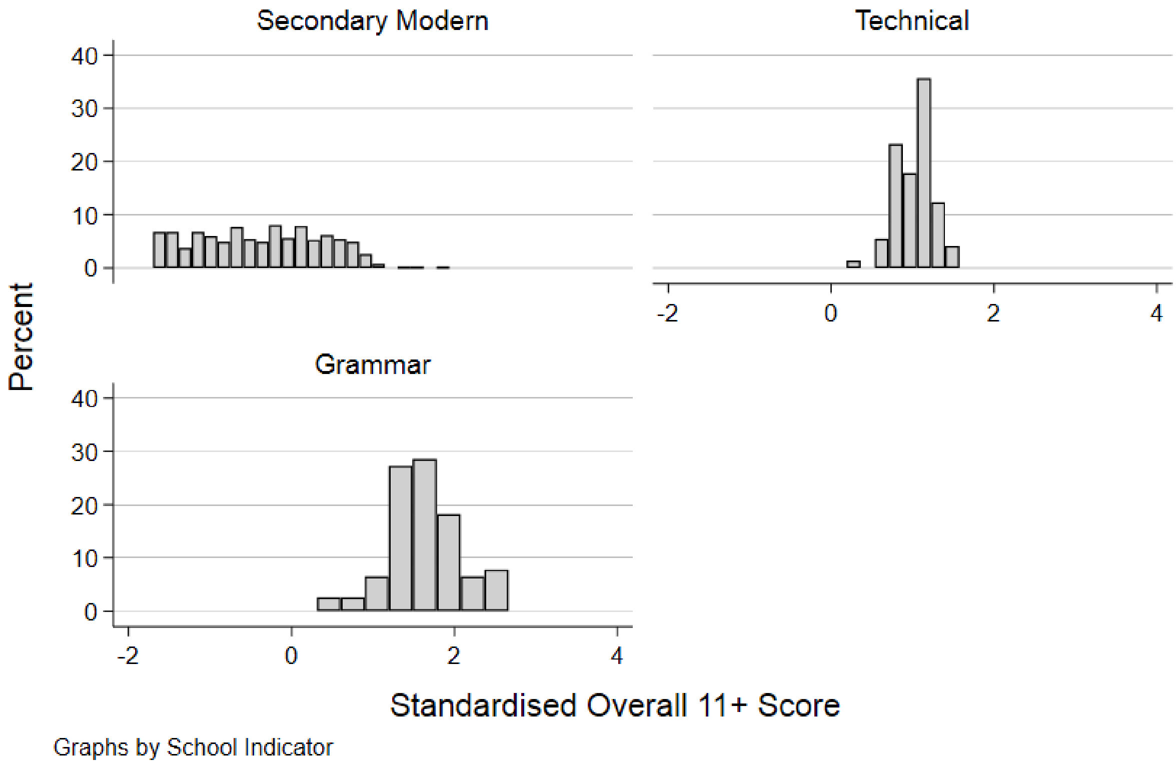


Figure 2-4 Distribution of standardised 11+ scores within each school type

Note: Average scores across the 11+ examinations separated by the school the child ultimately attended showing no clear cut-off but tendency for highest scorers to attend grammar school and lowest scorers to attend secondary moderns. See Appendix Table A2.1 for more detail.

School type was only recorded for 699 of the sample and is less detailed in terms of educational achievement than the continuous measure of 11+ scores previously mentioned. Children with typically developing speech and language profiles were approximately 2.5 times more likely to be selected into grammar school. Technical schools were relatively common within this cohort considering their rarity nationally and were more commonly attended by males than females; in 1948 the number of technical schools in the country peaked at just 318 (Bolton, 2012).

The study members sat further cognitive tests during high school. These were:

- The Raven's Progressive Matrices test (Raven, 1958) at ages 12 and 15, measuring IQ
- The Mill Hill Vocabulary Scale (Raven, 1958) at ages 12 and 15, measuring receptive vocabulary.

These two scales are reported to measure a child's capability to benefit from education (Miller et al., 1974; Raven, 1958). The four outcome measures (two scales at two time points) are consistent with the results presented for the 11+ examination with stronger effects of SLCN (and each subgroup) observed on vocabulary than on IQ. These results are not reported in the main text as they simply support the conclusions drawn from the 11+ measure.

2.8.1.2 Education Level and Formal Academic Qualifications by Adulthood

The highest educational attainment of the study member was recorded for those who returned to the study in adulthood (n=498). This was recorded as five indicators, "No Qualifications", "O-Levels or CSEs", "A Levels", "Degree" and "Post-Graduate Degree". As the differences in groups are categorical and not linear, this was modelled using an ordered logistic regression. The same covariates and SLCN measures reported in Table 2.7 are used as for the linear regressions. The ordered logistic estimated parameters were interpreted as Odds Ratios. This measure was also dichotomised into those with formal academic qualifications (n = 312) and those without qualifications (n = 186) and modelled using generalised least squares regression (ordinary least squares with heteroskedasticity-robust standard errors) in the same manner as all other outcomes.

2.8.2 Occupational Measures

2.8.2.1 Household Logged Income

Gross household income was recorded for each cohort member in the survey at age 50 and 62. This was recorded using categorical bands for annual income shown in Table 2.9.

For the current analysis, midpoint interval estimation is used. This is simply the average of the upper and lower limit used as the estimated income for each individual in the sample. For example, for the 18 individuals with household income between £40,000 and £45,000 at age 62, we estimate that their household income is £42,500.

As income is typically positively skewed (i.e. it has a long right hand tail in the distribution due to the small number of individuals earning very large amounts), the natural log of the midpoint estimate of the household income was taken and used as the outcome variable. Parameters estimated from the model can be interpreted as approximately the percentage change in income from a unit change in the predictor or identification variable. For example, a coefficient of 0.1 would represent an approximate 10% increase in household income.

Table 2-9 Distribution of household income categories at ages 50 and 62

Lower Limit	Upper Limit	Age 50	Age 62
£0	£2,500	8	1
£2,500	£5,000	40	9
£5,000	£10,000	47	39
£10,000	£15,000	69	45
£15,000	£20,000	75	44
£20,000	£25,000	55	48
£25,000	£30,000	60	33
£30,000	£35,000	30	30
£35,000	£40,000	20	18
£40,000	£45,000	10	8
£45,000	£50,000	10	6
£50,000	£60,000	8	9
£60,000	£70,000	4	3
£70,000	£80,000	1	2
£80,000	£90,000	0	0
£90,000	£100,000	0	3
£100,000	No upper limit	4	5
Total		**442**	**313**

Predicted interval estimation is where an individual's income is estimated by a regression with the estimate bounded by their recorded lower and upper limits. However, given that the relevant variables for generating this outcome variable would be included in either the final regression or as outcomes in other analyses, this would create endogeneity and induce correlation across multiple results. Furthermore, any regression for predicting income within the bounds results in heaping of individuals at the upper and lower limits and therefore does not add much continuous variation. Therefore, this method was not used in the final analyses.

The use of any interval estimate is limited by one major drawback in that it excludes the top category. The highest earners (>£100,000 p.a.) have no recorded upper limit to their income in the questionnaire and are therefore omitted from the analysis. Within this category there are two Individuals with SLCN at age 50 and one SLCN individual at age 62.

Two robustness checks were carried out to ensure that omission of this category did not drive results.

Firstly, a value of £105,000 was used to estimate the income of this group, in line with the progression of midpoint incomes across the prior categories. Findings are robust to this method.

Secondly, a Singh-Maddala distribution is applied to the interval distribution of income (Singh and Maddala, 1976). This gives a new estimate of income using maximum likelihood estimation for every category and is reliant upon three parameters; the number of participants per category; the lower limit of each category; the upper limit of each category with an upper limit. This fits an income distribution to the categorical distribution including a value of income for the top category at each time point and has been used in the British Cohort Study 1970 in the past (Blanden et al., 2007). This method was not used as the main analysis due to its complex interpretation and the small number of individuals in the top category (Gregg and Macmillan, 2010). The findings presented in the main results are robust to this method.

The household income measure at age 50 is the first in the literature to extend far enough to provide a reasonable measure of maximum earnings (Gregg et al., 2014; Grawe, 2006). However, the age 62 measure is unlikely to add significant value due to the high rate of retirement amongst the sample. Two thirds of respondents had retired, meaning that the income measure is less reflective of the maximum earnings potential of the individual or household.

For this cohort, employment was measured retrospectively on the questionnaire for ages 15, 25 and 35 by asking the participants to give their occupational status from the corresponding year, answered at age 50. This suggests that any unemployment recorded is likely to have been long term although this is not measurable. At age 50 and 62 they also filled in their current occupational status.

2.8.2.2 Employment

Employment was not tested as an outcome variable as only a very small percentage of the sample reported spending any time unemployed. Unemployment was asked retrospectively for 1962, 1972, 1982 and prospectively asked in 1997 and 2009. It is possible that the self-reporting of retrospective unemployment for a given year is subject to recall bias which may explain the consistently high employment records with little variation. The data available on unemployment is reported in Table 2.10 and appears heavily skewed by the presence of a very small number of individuals who were unemployed at multiple time points.

Table 2-10 Periods spent unemployed by typical and SLCN groups

Periods spent unemployed	'Typical'	SLCN
0	380	81
1	22	4
2	2	3
3	0	1
4+	0	0
	380	89

2.8.3 *Health Measures*

The Newcastle Thousand Families Study (1947) is an epidemiological study meaning that there is very detailed health data available. The literature discussed in Section 1.3.4 suggests a link with psychiatric risk, although little research has considered physical health outcomes.

2.8.3.1 General Health Questionnaire

The General Health Questionnaire contains 28 items measured on likert scales. This is a widely used tool for screening psycho-social wellbeing and identifying those at risk of psychiatric disorders (Goldberg and Williams, 1991). The GHQ-28 version is a shortened scale from the initial 60 item measure (Goldberg and Hillier, 1979) with four sub-scales: Somatic symptoms, Anxiety/Insomnia, Social Dysfunction and Severe Depression.

The combined scores are standardised (mean = 0, std. dev. = 1) for all participants within the sample at age 50 and at age 62. This score is reversed such that higher scores in the results indicate greater health. This allows for all results and outcomes to be comparable in terms of regression coefficients.

2.8.3.2 CASP-19

The CASP-19 scale is made up of 19 likert scale items measuring Control, Autonomy, Self-realisation and Pleasure (Hyde et al., 2003). This scale is targeted at measuring quality of life within older adults with a specific focus on happiness as opposed to measuring the absence of negative mental well-being. This provides a broader perspective of mental health than the GHQ-28 alone.

The scales are summed (range 0-57) and then standardised (mean = 0, std. dev. = 1) for all participants. A higher score represents a higher quality of life.

2.8.3.3 Grip Strength

Grip strength was measured three times for each participant who attended a clinical examination at age 62. The strongest score recorded of the three is used to generate the measure used for the analysis. This has a 0.99 linear correlation with the average grip strength score, demonstrating the stability of the measure. Grip strength is used generally as a crude measure of mortality within this age group, as discussed in the literature review.

The strongest grip recorded grip strength is predicted using sex and weight at age 62. The difference between the observed grip strength and the predicted grip strength is taken. This difference is standardised (mean = 0, std. dev. = 1) and analysed as grip strength adjusted for sex and weight.

2.8.4 Personality Measures

The OCEAN (a.k.a. the Big 5 or Five Factor Model) characteristics are the most widely accepted and reliably measurable foundation for non-cognitive skills. These are measured at age 62 within this dataset, using a highly replicable 60-item structure (reproduced from the NEO-FFI), where each trait has 12 items, measured on a 1-5 Likert scale. Each trait is standardised (mean = 0, std. dev. = 1) for the analysis.

The correlation matrix of the 5 personality traits ranged from 0.01-0.37. This demonstrates that each of the Big 5 traits are distinct from one another and is consistent with the literature (Heckman and Kautz, 2012). As discussed in greater detail within Section 1.3.3, the OCEAN characteristics are defined as follows.

2.8.4.1 Openness to Experience

"The tendency to be open to new aesthetic, cultural or intellectual experiences" (McCrae and Costa, 1992).

2.8.4.2 Conscientiousness

"The tendency to be organised, responsible and hardworking" (McCrae and Costa, 1992).

2.8.4.3 Extraversion

"The orientation of one's interests and energies outwards as characterised by positive affect and sociability" (McCrae and Costa, 1992).

2.8.4.4 Agreeableness

"The tendency to act in a cooperative, unselfish manner" (McCrae and Costa, 1992).

2.8.4.5 Emotional Stability

"A high level of predictability and consistency in emotional reactions and low susceptibility to rapid mood changes and psychological distress" (McCrae and Costa, 1992).

2.9 Analytic Strategy

The analysis presented will follow the structure depicted in Figures 2.5a to 2.5c, each figure representing a different measure of speech and language development as detailed above.

Each covariate is measured before the age of 7;6 in order to restrict the potential for the effect of speech and language development to be mediated by the covariates. The covariates are dummy, categorical or continuous variables. These are discussed first in the measures section.

The outcomes are measures from three key time points within the data collection:

- Age 11 (11+ scores used to determine schooling selection)
- Age 50 (First adult contact with full cohort)
- Age 62 (Most recent contact with full cohort)

There is gradual attrition over these three time periods as previously described.

Each analysis is ran using the participants with full information for that given outcome. Therefore, outcomes from different time points will have different available samples. Outcomes measured at the same time point will have nearly identical samples as the response rate per question of those who return the questionnaire was very high. This is detailed in the flow diagram of attrition (see Figure X).

The vast majority of the sample have full information for their speech and language development and every covariate. The primary source of missing data is for the outcome measures. For those missing on a given outcome, they are highly likely to be missing from that entire questionnaire, meaning that auxiliary variables (those correlated with the outcome but not included in the regression) are not identifiable. In the case where only the dependent variable is missing, and auxiliary variables are not identified, it is recommended that the complete case analysis proposed is conducted instead of single of multiple imputations (Jakobsen et al., 2017).

The format for the complete case analysis conducted according to the samples present in Figure 2.3 are shown in Figures 2.5a to 2.5c. Each analysis figure represents only a change in the measure of SLCN, utilising the 3 measures summarised in the Speech and Language Development Data: Section 2.7.6 Summary.

The ‘covariates’ referred to in Figures 2.5a to 2.5c are detailed in Table 2.7. The ‘outcomes’ in Figures 2.5a to 2.5c are summarised within the Measures: Outcomes section.

Generalised Least Squares (GLS) regression is used for all outcomes with one exception. This is equivalent to Ordinary Least Squares with Huber/White standard errors to correct for heteroskedasticity (Huber, 1967; White, 1982). Maximum Likelihood Estimation (an alternative estimator, commonly used in Structural Equation Modelling in the presence of missing data) are reported in Appendices (Appendix Figure A2.1) as a robustness check to this methodology.

In model (1), the coefficient represents the average change in outcome resulting from being in the ‘SLCN’ group as opposed to the ‘Typical’ group.

In model (2), the coefficient represents the average change in outcome resulting from being in the relevant SLCN subgroup as opposed to the ‘Typical’ group.

In model (3), the coefficient represents the average change in outcome resulting from one additional negative observation made by a health visitor.

For standardised outcomes, this average change is interpretable as the number of standard deviations of change in the outcome. For binary outcomes (Formal Academic Qualifications), this average change is interpreted as the change in the probability (range 0 to 1) of obtaining formal academic qualifications.

For the analysis of Education Level, which is an ordered categorical variable, ordered logistic regressions are used due to the non-linear progression between categories. In model (1), a regression coefficient with a value above one would suggest higher education levels within the SLCN group and a value below one would suggest lower education levels within the SLCN group. Similarly, in model (2), a regression coefficient above one for a given SLCN subgroup would indicate greater odds of achieving high education levels for that subgroup and a value below one would indicate the reverse. In model (3), the regression coefficient represents the odds of achieving high education levels for an individual with a given ‘Frequency’ of health visitor observations containing noted difficulties relative to an individual with one fewer concern noted. Hence, a value above one would indicate higher education levels within individuals with higher ‘Frequency’ and vice versa.

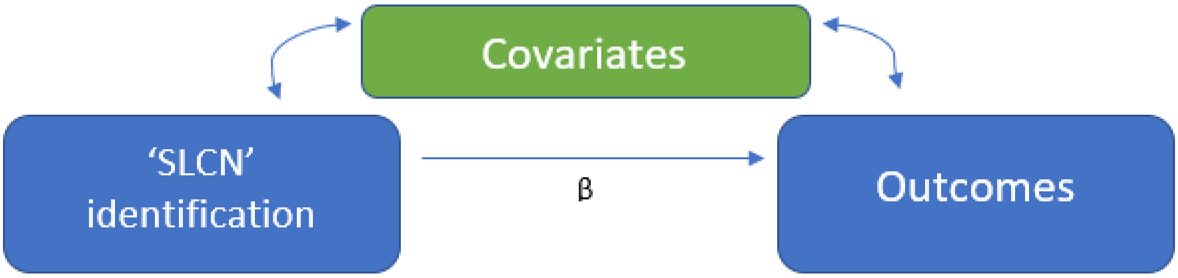


Figure 2-5a SLCN Model (1)

Note: β captures the estimated direct effect of an 'SLCN' identification, relative to a 'Typical' language profile, on a given outcome after controlling for a range of covariates.

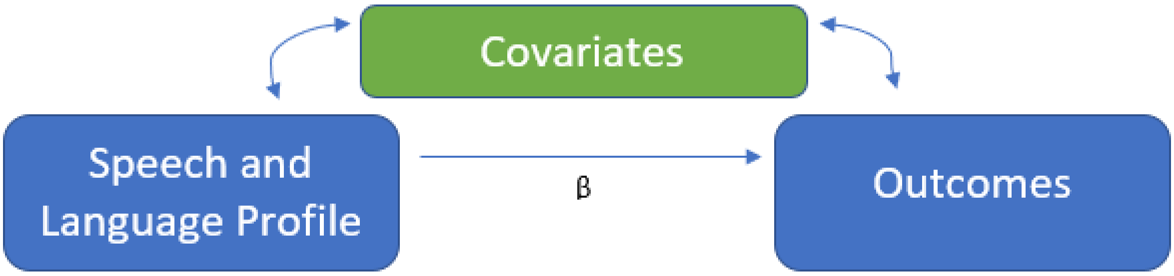


Figure 2-5b – Subgroup Profile Model (2)

Note: β captures the estimated direct effect of a specific SLCN subgroup (Resolving, Emergent or Persistent) identification, with one β per subgroup included simultaneously in the regression, on a given outcome after controlling for a range of covariates.

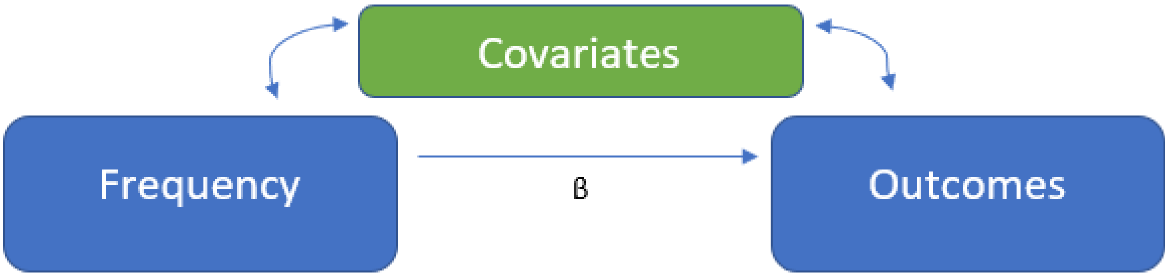


Figure 2-5c Frequency Model (3)

Note: β captures the estimated direct effect of each additional noted observation containing a concern from the health visitor about a child's speech and language development on a given outcome after controlling for a range of covariates.

2.10 Results

2.10.1 Education

2.10.1.1 Research Question

2a) What are the effects of SLCN on academic trajectories from the 11+ through to the maximum academic attainment observed by adulthood?

Early speech and language development have a statistically significant ($p<0.01$) predictive effect on 11+ scores within this sample. Children with SLCN score 0.23 standard deviations lower than the 'Typical' group on average after accounting for covariates. This effect is strongest within the 'Persistent' SLCN group. The coefficient remains negative for the other two SLCN subgroups although this coefficient is statistically non-significant, so we cannot confirm at the 90% confidence level that the 11+ score differences between our 'Resolving', 'Emergent' and 'Typical' groups are not zero.

Each of these results was consistent across each of the three sub-scales of the 11+ with the weakest effects of SLCN found on the arithmetic test ($\beta = 0.148$). The stronger correlations were found on the English and Verbal Reasoning sub-scales ($\beta = 0.256$*** and 0.267***, respectively).

Frequency was a strong predictor of the 11+ outcomes suggesting that the measure of persistence and severity was effective in identifying the potential for the child performing less proficiently in an academic examination.

Over time, you would expect the predictive capacity for SLCN to explain academic outcomes to decrease as other influences grow in relative importance. However, we still observe that the 'Persistent' SLCN group are around 20 percentage points less likely on average to obtain any formal academic qualifications than the 'Typical' reference group.

Table 2-11 Education results

Note: Estimates are reported with robust standard errors below in parentheses. A negative estimate indicates poorer expected outcomes for the group identified by that row for the first two outcomes. An odds ratio of below one indicates poorer expected outcomes for the group identified by that row. * = p < 0.1, ** = p < 0.05, *** = p < 0.01

	Standardised 11+ Score			**Any Formal Academic Qualifications**			**Education Level (1-5, Adjusted Odds Ratios)**		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
SLCN	-0.233***			-0.084			0.641*		
	(0.090)			*(0.059)*			*(0.165)*		
Resolving		-0.096			-0.032			0.903	
		(0.138)			*(0.085)*			*(0.324)*	
Emergent		-0.178			0.238			1.62	
		(0.224)			*(0.162)*			*(1.03)*	
Persistent		-0.349***			-0.200**			0.336***	
		(0.120)			*(0.079)*			*(0.129)*	
Frequency			-0.030***			-0.010*			0.943**
			(0.010)			*(0.006)*			*(0.026)*
N	**712**	**712**	**712**	**441**	**441**	**441**	**441**	**441**	**441**
(pseudo) R^2	**0.203**	**0.206**	**0.207**	**0.185**	**0.198**	**0.186**	**(0.104)**	**(0.110)**	**(0.106)**

However, the historical context is important for interpretation here, as discussed in the literature review. Formal academic examinations at school leaving age were not compulsory in this era and many of the participants without formal academic qualifications will have chosen to leave school before sitting these exams and therefore are unlikely to have failed on account of never having entered. Given the prevalence of trade jobs, demonstrated by the parental occupational class distribution, formal academic qualifications may not have been considered necessary by many of the participants. Without knowing the reason for the lack of formal academic qualifications obtained by age 50, it is hard to judge whether this result would be indicative of future labour market difficulties in the same way as in the modern context.

The final academic measure used is the highest education level obtained, taken as a categorical ordered variable. The coefficients reported represent odds ratios relative to the 'Typical' group. The SLCN group were 66% as likely to have achieved each additional level of education, or in other words they were a third less likely to have made each progression. This odds ratio of 0.66, adjusted for all covariates, is "educationally significant" according to the cut-offs used in Lindsay and Strand (2016) of 0.75 or lower (equivalent of the 1.33 or higher cut-off when measuring a negative outcome).

The histogram shown in Figure 2.6 demonstrates that this is not due to a particular difficulty at any given level of academia but that the SLCN group are less likely in general to proceed further in education at each given stage.

Again, this effect is more pronounced for the 'Persistent' SLCN group also depicted in Figure 2.6. The 'Persistent' SLCN group were approximately three times less likely to progress through the categories than the 'Typical' group. Figure 2.6 clearly shows the high probability that an individual identified with 'Persistent' SLCN in childhood will leave compulsory schooling without any formal academic qualifications. This effect is visibly pronounced relative to the 'Resolving' SLCN subgroup for whom academic progression appears relatively similar to the 'Typical' group.

As there are only eight remaining participants from the 'Emergent' SLCN subgroup, interpretation of their results is generally limited as it is highly susceptible to the influence of outliers.

Whilst persistent and frequent forms of SLCN were significantly associated with the education level achieved by the individual, these factors were far less predictive of this outcome than either social class or gender.

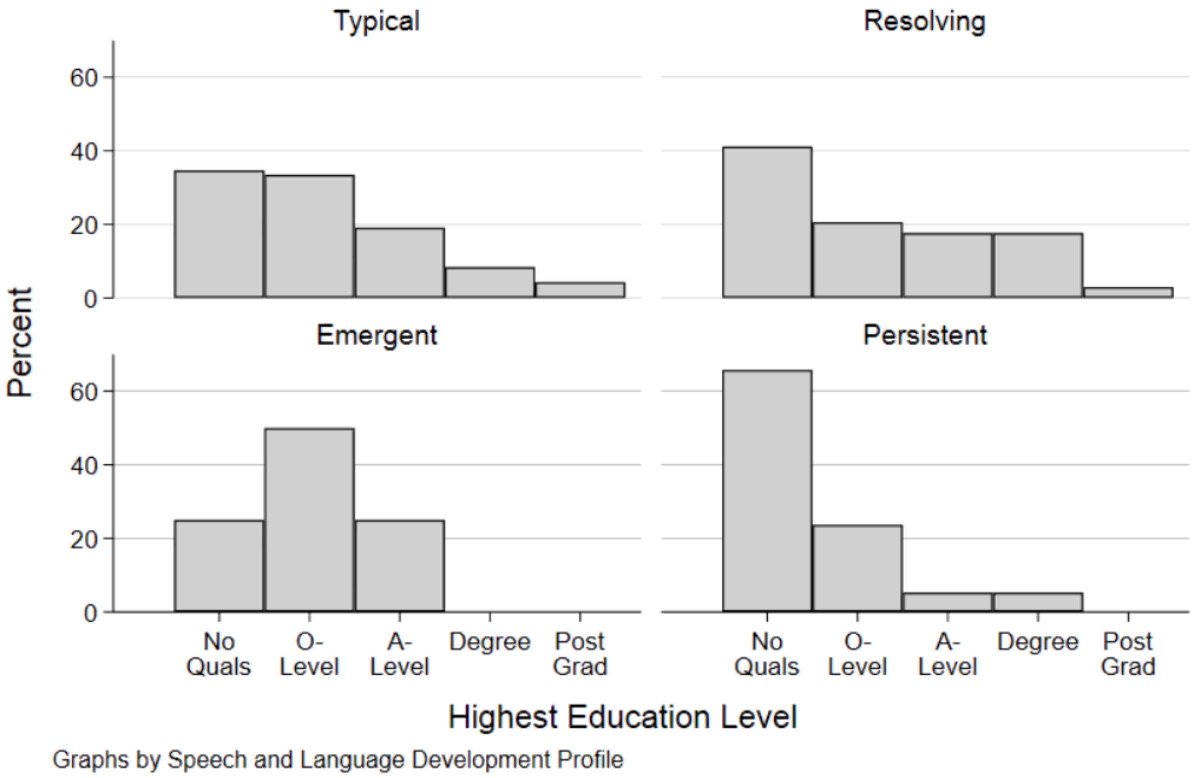


Figure 2-6 Highest academic qualification distributions by SLCN profile

There were further cognitive tests measuring an individual's capacity to benefit from education taken at ages 12 and 14, as mentioned in the Measures section. These were the Raven's Progressive Matrices test and the Mill Hill Vocabulary Scale, each taken at both ages. The results of the equivalent analyses on these measures were entirely consistent with the results shown from the 11+ score tests and so these were not reported for brevity.

The estimated coefficients suggested that the 'Persistent' SLCN group score between 0.25 and 0.5 standard deviations lower than the 'Typical' group on these cognitive tests. This is consistent with the results presented in Table 7 for the average of the 11+ test scores and suggests that the measured group differences at age 11 were maintained through early adolescence.

2.10.1.2 Summary

Early speech and language development significantly predict later educational performance.

'Typical' development predicts higher scores at age 11 which were crucial at the time for determining which type of school the child would attend. 'Typical' development also predicts a higher probability of obtaining formal academic qualifications and progressing into each stage of further education.

Within the SLCN subgroups, these findings were most pronounced for the 'Persistent' SLCN across all measures. Similarly, the more frequently the child displayed difficulties with their communication development to the health visitor, the lower the expected academic attainment for each outcome.

2.10.2 Occupational

Table 2-12 Occupational (income) results

	Logged Household Income at 50			**Logged Household Income at 62**		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
SLCN	-0.168*			-0.073		
	(0.100)			*(0.101)*		
Resolving		-0.179			-0.202	
		(0.147)			*(0.132)*	
Emergent		-0.019			0.156	
		(0.272)			*(0.322)*	
Persistent		-0.191			0.049	
		(0.132)			*(0.148)*	
Frequency			-0.025***			-0.001
			(0.010)			*(0.011)*
N	**438**	**438**	**438**	**308**	**308**	**308**
R^2	**0.115**	**0.115**	**0.122**	**0.112**	**0.120**	**0.111**

* = p < 0.1, ** = p < 0.05, *** = p < 0.01

2.10.2.1 Research Question

2b) What are the effects of SLCN on earnings in adulthood?

There is mixed evidence of whether a link between childhood speech and language development and later life earnings exists.

At age 50, the average household income of the participants identified with SLCN was approximately 16.8% lower than the average household income of the participants identified as ‘Typical’. This is statistically significant at the 10% level but not the 5% level.

However, given the reliance on midpoint estimation within a categorical income band and the fact that the income is measured on a household basis, this is a strong finding.

Furthermore, a 16.8% difference in income is substantial to the quality of life within a household. The 95% confidence interval ranges from -2.8% to +36.4%.

Each subgroup of SLCN shows a non-significant negative association with income at age 50 although this is very weak for the small ‘Emergent’ profile.

Notably, the ‘Frequency’ variable, indicating the persistence of SLCN as a proxy for severity, shows a significant relationship with income at the 1% significance level.

At age 62, the same analysis is conducted. However, all association was no longer present at this age. Two thirds (205 of 308) of the sample had retired in the interceding 12 years which makes our household income measure a poor reflection of their earnings potential as a household or individual. The estimates produced at the two timepoints are notably different, limiting the reliability of the association found. Whilst the age 50 measure is likely to be more reliable, a single timepoint is susceptible to temporal fluctuations in employment, as discussed in the literature review.

2.10.2.2 Summary

There are potentially important group differences in earnings between the SLCN and Typical groups. However, the measure at age 62 is likely to be a poor reflection of the participants’ productivity or lifetime earnings, whilst the measure at age 50 only gives statistically weak, but economically large, results.

Our estimate suggests that, after controlling for covariates, the SLCN group may earn -3 to +36% less than the Typical group as a household on average. More research would need to be conducted in an alternative data set to satisfactorily answer the research question. Furthermore, a more accurate estimate of earnings differences driven by early speech and language development data would allow for the testing of potential pathways of interest. In particular, performance in the education system is likely to be an important predictor of earnings potential but this data is limited in its potential to test any hypothesis.

2.10.3 Health

Table 2-13 Health and quality of life results

	Subjective Health (GHQ-28) at 50			Subjective Health (GHQ-28) at 62		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
SLCN	-0.273**			0.079		
	(0.135)			*(0.155)*		
Resolving		-0.388**			-0.055	
		(0.196)			*(0.216)*	
Emergent		-0.292			0.546	
		(0.358)			*(0.500)*	
Persistent		-0.156			0.320	
		(0.180)			*(0.215)*	
Frequency			-0.019			0.014
			(0.013)			*(0.016)*
N	**425**	**425**	**425**	**308**	**308**	**308**
R^2	**0.062**	**0.063**	**0.060**	**0.090**	**0.100**	**0.091**

	Quality of Life (CASP-19 Scale) at 62			Grip Strength at 62		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
SLCN	0.092			0.060		
	(0.158)			*(0.162)*		
Resolving		0.093			-0.043	
		(0.214)			*(0.221)*	
Emergent		-0.067			0.146	
		(0.512)			*(0.480)*	
Persistent		0.120			0.161	
		(0.225)			*(0.238)*	
Frequency			0.002			-0.020
			(0.017)			*(0.016)*
N	**320**	**320**	**320**	**277**	**277**	**277**
R^2	**0.083**	**0.083**	**0.082**	**0.106**	**0.108**	**0.111**

* = p < 0.1, ** = p < 0.05, *** = p < 0.01

2.10.3.1 Research Question

2c) What are the effects of SLCN on health in adulthood?

There is very little evidence that the early speech and language development of an individual directly predicts their health outcomes at age 62, whether subjective, objective or quality of life based.

There is some evidence that at age 50, there are group differences in health outcomes. The SLCN group report health difficulties at a rate of around 0.27 standard deviations more regularly than the 'Typical' group (GHQ is a questionnaire noting health difficulties which was recoded to measure health positively in order to be comparable to all other outcome measures).

The GHQ-28 is made up of four sub-scales. Analysis of these sub-scales show that the SLCN group only reported greater difficulties on the severe depression sub-scale at age 50 and not on the 'social dysfunction', 'anxiety and insomnia' or 'somatic symptoms' scales. However, this is primarily driven by the 'Resolving' SLCN profile and there is no theoretical reason why this group would be more likely to report health problems than the 'Persistent' SLCN profile for example. Similarly, there is no theoretical reason why the SLCN group would report difficulties on the severe depression sub-scale but not the social dysfunction or anxiety sub-scales which are more clearly associated with known precursors identified in the literature (see literature review for more detail).

The GHQ-28 scale has an identified clinical cut-off used to indicate higher risk of developing psychiatric disorders. This cut-off was applied to the age 50 and age 62 continuous scores to generate a dummy variable. T-tests reveal no significant differences between the 'Typical' and SLCN groups ($p = 0.214$ and $p = 0.216$) respectively.

The SLCN group report statistically equivalent levels of quality of life and grip strength at age 62 to the Typical group.

2.10.3.2 Summary

Early speech and language development did not predict health at age 62 within this cohort. There were some inconsistent negative associations between health at age 50 and 'Resolving' SLCN.

Overall, there is very limited evidence that early speech and language development data was a precursor to later life health problems within this cohort.

2.10.4 Non-Cognitive Skills

2.10.4.1 Research Question

2d) What are the effects of SLCN on the OCEAN non-cognitive skills as measured in adulthood?

As shown in Table 10 on the next page, there are differences in some of the Big 5 personality traits for SLCN participants although other traits show no differences at all. In particular, Openness to Experience and Extraversion are significantly lower within the SLCN participants after accounting for covariates. Conscientiousness, Agreeableness and Emotional Stability are not significantly different between the two groups. See the Glossary section for full definitions of each of the Big 5 non-cognitive skills.

Openness to Experience is significantly lower in the SLCN subgroup, particularly for the 'Persistent' SLCN group who are estimated as almost 0.7 standard deviations lower than the 'Typical' group after accounting for covariates. Openness to Experience measures an individual's enjoyment in seeking out next experiences and is strongly linked to their potential to succeed and pursue academic endeavours. This relationship is endogenous and recursive, however. As we don't observe this trait until age 62, we cannot state here whether early difficulties with school cause lower Openness to Experience, or whether lower enjoyment or learning causes lower probability of performing well in an educational setting (or whether the correlations are totally spurious).

Conscientiousness shows no significant differences between the Typical and the SLCN group(s). This encompasses traits such as work ethic, attention to detail and organisational skills within adults. The literature review did not reveal any common connotations between these skills and SLCN groups. However, the childhood analogous traits (see the literature review discussion of John and Sririvista, 1999) of impulse control, attention span and delay of gratification do have connotations to the behaviour problems which are common in children with SLCN. It is therefore of interest that the measure of Conscientiousness used here at age 62 does not show any association with SLCN identification. These results suggest that if any differences exist between children on their early conscientiousness related to communication development, they are no longer present in adulthood.

Table 2-14a Non-Cognitive Results

	Openness to Experience			Conscientiousness			Extraversion		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
SLCN	-0.340**			-0.077			-0.374**		
	(0.163)			*(0.160)*			*(0.158)*		
Resolving		-0.023			0.132			-0.343	
		(0.216)			*(0.218)*			*(0.211)*	
Emergent		-0.763			-0.277			-1.164**	
		(0.517)			*(0.512)*			*(0.588)*	
Persistent		-0.684***			-0.019			-0.301	
		(0.231)			*(0.225)*			*(0.225)*	
Frequency			-0.030*			-0.006			-0.016
			(0.017)			*(0.017)*			*(0.017)*
N	**310**	**310**	**310**	**306**	**306**	**306**	**310**	**310**	**310**
R^2	**0.057**	**0.077**	**0.053**	**0.054**	**0.055**	**0.054**	**0.051**	**0.058**	**0.0.37**

* = p < 0.1, ** = p < 0.05, *** = p < 0.01

Table 2-14b Non-cognitive results (continued)

	Agreeableness			Emotional Stability (- Neuroticism)		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
SLCN	-0.086			0.066		
	(0.151)			*(0.152)*		
Resolving		-0.108			-0.081	
		(0.206)			*(0.204)*	
Emergent		0.507			0.441	
		(0.483)			*(0.498)*	
Persistent		-0.167			-0.136	
		(0.212)			*(0.218)*	
Frequency			-0.001			-0.004
			(0.016)			*(0.016)*
N	**316**	**316**	**316**	**316**	**316**	**316**
R^2	**0.082**	**0.087**	**0.081**	**0.111**	**0.112**	**0.109**

* = p < 0.1, ** = p < 0.05, *** = p < 0.01

Extraversion is significantly lower amongst individuals with SLCN. This broad non-cognitive skill encompasses traits such as sociability and assertiveness. Whilst there is evidence of externalised behaviour difficulties within SLCN children (see literature review, particularly Beitchman et al., 2001), there is also the potential for withdrawal within children who struggle to communicate (see literature review, particularly Dockrell et al., 2014). Therefore, this result is consistent with our understanding of the expected personality development of children and adolescents who struggle to communicate. The fact that such strong personality associations remain at age 62 with speech and language development measures ceased at age 7;6 is clear evidence of the important role that early speech and language development can play in so many aspects of an individual's life.

There are no significant differences between participants in the SLCN and the Typical groups on their measure of Agreeableness at age 62. Agreeableness contains traits related to prosociality, including politeness and co-operation (Zhao et al., 2016). Prosociality has been previously shown to be moderate to high amongst individuals with Specific Language Impairment, or latterly Developmental Language Disorder (Conti-Ramsden and Durkin, 2016; Toseeb et al., 2020) although this research does not state that language impaired individuals are specifically more prosocial than typically developing groups, just that they are not less prosocial. There is no effect detected in the current study and this appears consistent with the literature at the time of writing although this literature is rapidly growing.

Emotional Stability similarly shows no significant association with identification of SLCN. The literature on behaviour within SLCN groups suggested a possible indication of lower Emotional Stability. However, we only have a measure of the adult facets of emotional stability at age 62 for which no SLCN literature previously existed. By age 62, within our sample, there are no group differences between SLCN and Typical groups for their measured Emotional Stability.

2.10.4.2 Summary

There are significantly lower levels of Openness to Experience and Extraversion within the people with a history of SLCN. These are the two strongest effect sizes estimated in the entire analysis of the Newcastle Thousand Families Study. There was a small effect size[2] between SLCN and both Openness to Experience and Extraversion.

As discussed in the literature review, Openness to Experience and Conscientiousness are commonly found to be the most important of the Big 5 for predicting academic and labour market success. Longitudinal modelling of the development of Openness to Experience alongside academic achievement within comparable Typical and SLCN populations would be beneficial to unpicking this relationship. Fostering greater openness to new experiences indicates greater engagement in formal education although the direction of this effect is impossible to ascertain using only the data available. Understanding the relationship between educational achievement, openness to new experiences and the presence of SLCN may help to develop positive practices for improving outcomes for the SLCN group in academic settings.

The predictive power of the model is relatively small with only a low R^2 value for each regression. The separation of SLCN into subgroups tended to produce the most predictive models, particularly for Openness to Experience, suggesting that there is a notable difference between 'Resolving' and 'Persistent' SLCN difficulties.

[2] 0.5 > cohen's d > 0.2 (Cohen, 1988)
Cohen's d = 0.340 and 0.374 for Openness to Experience and Extraversion respectively.

2.10.5 Conclusions

The results presented in Tables 2.11-2.14 demonstrate a complex picture of the outcomes for children identified SLCN in early childhood.

Figure 2.7 demonstrates the comparable effects of SLCN status on each outcome, grouped by the sphere of life within which that outcome was grouped in our results, i.e., Education, Occupational, Health and Personality. To construct Figure 2.7, each outcome was standardised, including the income variable and education level indicator, in order to allow for a comparable X axis. The X axis represents the mean outcome for the 'Typical' group subtracted from the mean outcome for the 'SLCN' group on a standardised outcome, meaning that they plot the coefficient of our standardised regression output.

The strongest and most significant effects are found on the participants' educational attainment, Extraversion and Openness to Experience with each of these showing a significant disadvantage for the SLCN group. There are weak associations with both income and general health questionnaire scores at age 50. Further research would be necessary to satisfactorily clarify whether SLCN does predict lower earnings and a higher tendency towards psychiatric risk at this age as these effects are not found at age 62.

The educational correlations are particularly strong within the 'Persistent' SLCN group, i.e. the children who had noted difficulties with communication within their school years tended to perform worse across their subsequent academic careers. This is also true for the results relating to Openness to Experience, a personality trait which is heavily linked to engagement in education.

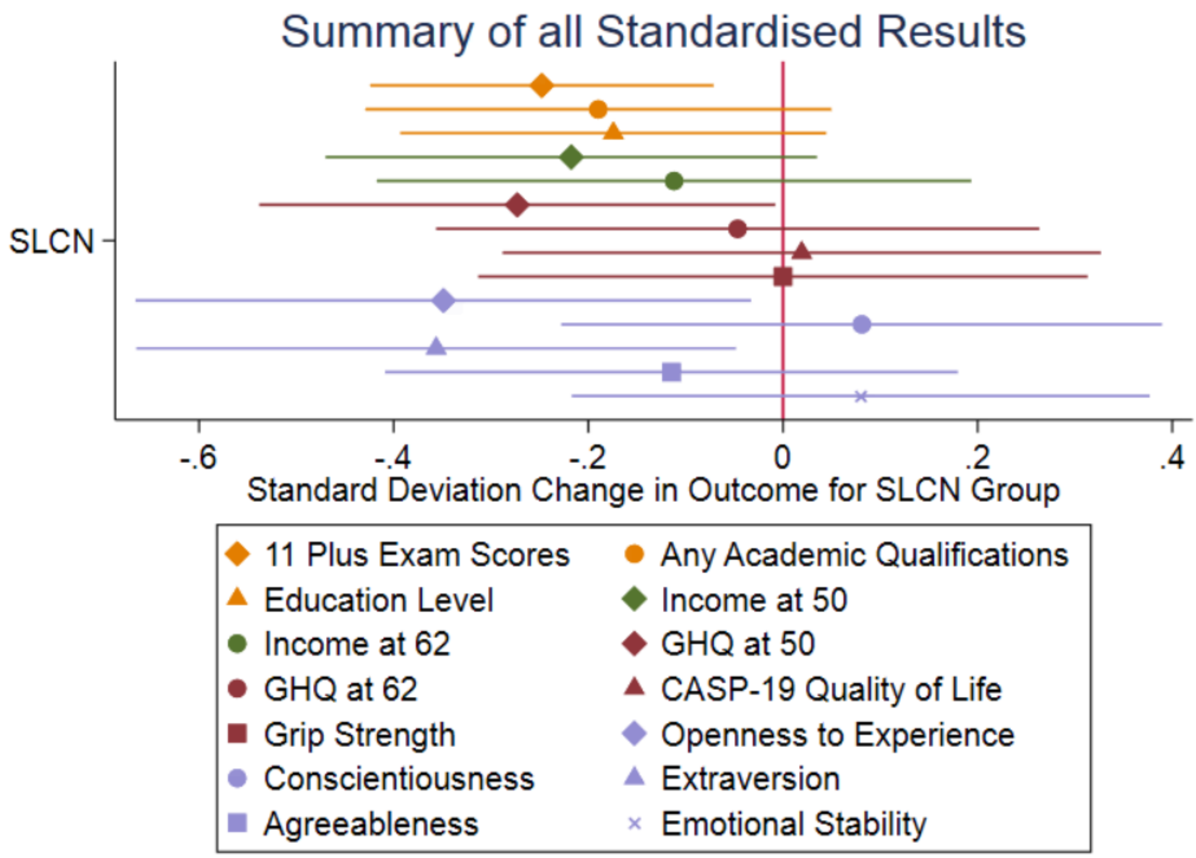


Figure 2-7 Standardised effect sizes from SLCN to all outcomes

Note: The X axis represents the estimated standardised effect sizes across each outcome. Point estimates are marked with horizontal lines representing a 95% confidence interval. Where this confidence interval crosses the vertical line at 0, the result is not significant. Where the confidence interval lies entirely to the left of the 0 line, this indicates significantly poorer outcomes for the SLCN group.

Note: Yellow = Education. Green = Occupational. Red = Health. Blue = Non-Cognitive Skills.

2.11 Discussion

2.11.1 Newcastle Thousand Families Study

The analysis and data presented is unique. This study represents the oldest population sample records of speech and language development in the literature with adult follow up. This data gives novel insight into the details of early speech and language development in the post-war era and how it was assessed by a leading speech therapist of the time.

We find that speech and language profiles were evident within the data and that they were largely consistent with modern estimates of prevalence at the relevant ages. However, direct comparability was not possible due to the unique nature of the data source as discussed in the Speech and Language Development Data Summary. The observed prevalence was relatively high although there were many factors which explain this previously discussed. Furthermore, the inclusive criteria had two major benefits: 1) Despite the later attrition in the sample, there are still enough people with a history of SLCN to allow analysis, including analysis of three subgroups and 2) coefficients estimated using a relatively mild definition of SLCN are unlikely to be overestimates of the effect of SLCN, increasing the confidence attributable to the significant findings.

It was also not possible to separate specific forms of SLCN due to the inconsistent recording of children with no standardised scales. Even within the four distinct profiles there is no homogeneity as speech sound disorders are grouped with those with low vocabularies, lisps, stammers and other forms of SLCN. Whilst there is substantial overlap in language disorders and speech disorders, particularly in the risk factors for persistent forms of both, the inability to distinguish forms of communication difficulty limits the potential for theoretical interpretations on the reasons behind later differences. For example, the tendency towards more introverted personality traits might imply reticence to vocalise thoughts in a child with a speech disorder, whereas it might imply more general exclusion from the classroom in a child with a receptive vocabulary disorder. The reasons why these traits might persist across the life course is similarly complex even if we could identify the specific difficulty. The measure used is clearly unique, and not equivalent to a formal diagnosis of SLCN or any other Special Educational Need. As this is a limitation of any historical measure, the following analyses chapters will utilise a continuous measure of vocabulary to acknowledge and explore the limited potential to identify specific groups.

It was possible, however, to separate SLCN groups according to the timing of their difficulties using the school entry date. This is commonly used in both historical and modern literature for speech and language sciences due to the significant effect that attending school has on a child's communication environment, motivation, and production.

It is worth noting that many of the children within this cohort received speech therapy. In fact, the high number of health visitor observations and direct contact with a speech therapist means that it is unlikely that many speech and language difficulties went unnoticed relative to a typical population. The intervention received was not recorded over time but may potentially mitigate the effects we would otherwise observe in a typical population.

The dataset itself is unique, representing an early baby boomer cohort within a specific city of the UK. Whilst the complete sample within two months of birth gives a strong population approach, the population itself remains relatively specific. The principles of early communication are unlikely to be dramatically different geographically, but the impacts this has on life chances may be more variable. For example, the job opportunities in Newcastle-Upon-Tyne over the latter half of the 20th century may not accurately reflect the national labour market of the time or the modern day. This limits the extent to which specific figures (e.g., the coefficient attributed to differences in income) can be used in other contexts. However, the broad correlations between variables, as shown in Figure 2.7, are likely to be a reliable first indication of how early speech and language development relates to outcomes in later adulthood.

Longitudinal data is essential to be able to test the long-term implications of early disadvantage. In the context of SLCN, the few studies with outcomes drawn from adulthood are commonly cited as evidence of the life-long importance of early speech and language development. This data source allowed for a broader scope when considering the potential long-term costs of childhood SLCN including two key political outcomes, earnings and health, and two key potential pathways to improved life chances through cognitive skills (education outcomes) and non-cognitive personality traits.

The overall picture is complex and there is not a simple finding across the spectrum of outcomes to summarise the "effect" of SLCN. There were, however, some clear findings which have implications for future research.

2.11.2 Educational engagement

It was well known that educational outcomes are poorer for people with SLCN. This research extends this finding to cover five levels of educational attainment achievable by adulthood and supports the Durkin et al. (2009) finding that those with persistent SLCN will be the most disadvantaged. Prior to even the regression results, the group differences in academic attainment into adulthood are clearly visible in Figure 2.6.

There are three important considerations to these findings presented below.

Firstly, the educational context has changed significantly from this cohort to the modern day. Examinations at age 11 are no longer the predominant method for selection of secondary schooling institutions. If selective schooling entrenches initial differences in academic attainment, then this move may be beneficial to individuals with SLCN who were more likely to attend a secondary modern school as opposed to a grammar school. However, associations with IQ exams at age 12 and 14 were consistent with those observed at age 11, suggesting that the gap between SLCN and typically developing individuals is not widened after the selective schooling process. Whether or not a comprehensive system would have resulted in more equal outcomes, either in terms of IQ or exam results, is not testable.

Examinations are now compulsory at the end of secondary education which drastically reduced the number of adolescents who leave school with no formal academic qualifications. Further education is far more commonly taken within modern cohorts than within this data. Whether this change has been equivalent across typical and SLCN groups is unclear and is worthy of future research.

Secondly, education is commonly seen as the vehicle for social mobility. As there is an established social gradient to SLCN, both in the literature and in this cohort, it is important to consider whether socioeconomic status could be interacting with SLCN status in its relationship to education outcomes. There are potentially different effects for different groups to consider. For example, would SLCN predict poorer academic attainment as strongly within high SES families as low SES families? There are potential risk and protective pathways into adulthood at play which would benefit from greater research. Within this cohort, there was only a limited measure of SES (father's occupational social class) and a clustered distribution within that (many fathers in Newcastle in 1947 worked skilled jobs and very few were at the upper and lower ends of the distribution). Both issues would be unlikely to be the case in more modern cohorts with a greater appreciation of the many aspects of socioeconomic status and a more polarised distribution of labour.

Thirdly, there is an established literature on the importance of Openness to Experience in educational attainment and it is unlikely to be coincidental that the SLCN group in this cohort experience lower levels of both. However, as Openness to Experience is measured only at age 62, we cannot model which factor is most important for driving the other. If, for example, higher Openness to Experience was the key driver of educational attainment (i.e. children who enjoy learning new things perform well in school and are more likely to continue into further education), then this would have significant implications for policy and targets for children with SLCN and teachers. If, in reverse, achievements in education fostered a greater exposure and enjoyment of learning (attendance at university, for example, may not simply increase exposure to more than academic information but exposure to other cultures and increase other facets of Openness to Experience), then there would be no need for a major change from the current focus on academic attainment. Within the SLCN literature, traits such as social withdrawal and behavioural problems are commonly observed and suggest limited engagement with education but it is unclear from this data whether Openness to Experience was reflective of these issues in the SLCN sample or at what age these differences develop.

2.11.3 Income

Household income is the measure captured in the Newcastle Thousand Families Study at age 50 and 62. There are positives to using household income in that it is a good reflection of household resources and the life of the participant. However, it also has drawbacks in that it is less likely to be a reflection of the individual's earning potential which we want to measure in order to test for any association with SLCN.

Income is often measured in categorical bands, as in this study, although this is still a limitation. A large amount of the variation in earnings is lost within these bands. The imputed values from the Singh-Maddala distribution similarly only attribute one estimated household income per category. This limitation is unlikely to be resolved in future research as it is rare for a cohort to ask for a more detailed measure of income than this.

However, the two income timepoints within the study reported mixed evidence of a link between SLCN and income. At age 50, incomes were approximately 17% lower for the households where the participant was identified with SLCN. The frequency of observed speech and language difficulties in childhood predicted household incomes at age 50 at the 1% significance level. By age 62, however, there was no significant association between SLCN and household income by the same measure.

The research question was therefore not robustly addressed. The high rate of retirement amongst respondents at age 62 meant that the income measure at this time was unreliable as an indicator of labour market success. Therefore, our finding of lower incomes within the SLCN group was susceptible to temporal mismeasurement as only one time point could be reliably compared across the sample. This single timepoint may easily have fallen on a non-representative time point in any individual's career. For example, if a participant was temporarily unemployed at the time of measurement, our estimate of their career success could be seriously flawed for extrapolation. This limitation is significant as the study aims to make conclusions about the effects of SLCN on income across the life course implied by the effects measured at age 50.

The evidence provided simply doesn't answer the research question satisfactorily. As this is a crucial research question to the aims of the thesis (to establish the long term implications of SLCN with a particular focus on policy-relevant costs including tax revenue), the next analyses exploit an alternate cohort study with a greater focus on occupational success.

2.12 Summary

The analysis presented suggests five key findings, relevant to the initial research questions:

- Early speech and language development profiles were observable within the data. The prevalence was comparable to the modern literature. The profiles were highly varied as specific forms of SLCN were not identifiable.
- Educational attainment was significantly poorer for children with SLCN within this cohort. This was evidenced from entry to secondary school throughout further education. These results were primarily driven by children with SLCN which persisted beyond school entry.
- It is unclear whether children with SLCN are likely to earn less in adulthood from the results presented. Further research is warranted and would ideally contain multiple

measures of income from at least age 30, to establish a stable estimate of maximum earnings capacity. However, estimating income at age 62 was found to be significantly restricted by the high proportion of the sample who had already retired.

- Health outcomes for children with SLCN appear relatively equal to children without SLCN. This contradicts prior research findings using alternative measures of mental health and this question would benefit from further research.
- Openness to Experience and Extraversion in late adulthood are significantly lower within children with SLCN. Conscientiousness, Agreeableness and Emotional Stability are not significantly different. The timing of these associations suggests a profound difference between the SLCN and the 'Typical' group to maintain differences in personality across 60 years of life. Research into adolescent personality traits suggests that these complexities are common and further research into these traits should be a key focus of the future of the field of speech and language sciences.

To further address some of these findings and limitations, the following analyses are conducted within a large, nationally representative sample with extensive longitudinal follow-up. In particular, a sample is required which contains multiple observations of income across adulthood, including timepoints which reflect maximum earnings. For these reasons, we now draw on data from the British Cohort Study (1970) to further assess the life course effects of childhood SLCN.

The following analysis chapter models earnings in adulthood using a standardised, widely used measure of early language development known as the English Picture Vocabulary Test (Brimer and Dunn, 1962).

3 Mediation Analysis of Earnings in the British Cohort Study 1970

3.1 Introduction

This chapter further interrogates the research question outlined, but not fully answered, in the previous chapter – Do early difficulties with communication predict lower earnings in adulthood?

Much of an individual's life is spent working. Earnings represent a crucial element in an individual's life experience, influencing their quality of life whilst reflecting the value of skills learned, demonstrated and applied in the labour market.

Two potential pathways that may help to explain any relationship with earnings are tested in a mediation analysis: education level and non-cognitive skills.

Chapter 1 summarises a significant body of evidence linking early vocabulary development to educational attainment and hypothesising relationships with the OCEAN non-cognitive skills. Wide-ranging evidence that education and non-cognitive skills are rewarded in the labour market is also presented (Griliches and Mason, 1972; Dolton and Vignoles, 2000; Heckman and Kautz, 2012).

Chapter 2 established clear differences in educational attainment at the completion of formal education, as well as lower levels of Openness to Experience and Extraversion at age 62, within the Newcastle Thousand Families Study for a group identified with Speech, Language and Communication Needs (SLCN) after accounting for a range of covariates (see Section 2.7 for a covariate list and Figure 2.7 for a summary of all results).

Given the noted rank stability of adult non-cognitive skills (O'Connell and Sheikh, 2010; Heckman and Kautz 2012) and the literature on early behaviour (van Daal et al., 2007), associations with non-cognitive skills are likely to be formed before adulthood. This could help explain some portion of any income gap between an identifiable SLCN group and a Typically Developing reference group if there is an income gap.

Openness to Experience has been commonly linked to educational outcomes in the literature (Heckman and Kautz, 2012) which may suggest that the two pathways are complementary in their contribution to observed income differences unexplained by a direct communication pathway.

This chapter intends to complement this initial research by utilising the British Cohort Study 1970. The following research questions are addressed in this Chapter.

3.2 Research Questions

1. Does a child's vocabulary impact their earnings in adulthood?
2. Does performance in the education system mediate any relationship between early vocabulary and earnings?
3. Are non-cognitive traits developed in later childhood mediators of any relationship between early vocabulary and earnings?

The following section outlines the data selected to answer these research questions. The measures of language, earnings, control variables and mediators are addressed in turn.

3.3 The British Cohort Study 1970

3.3.1 Recruitment, Data Collection and Attrition

The British Cohort Study 1970 is a nationally representative longitudinal sample of 17,827 children born between 5-11 April 1970 in the UK.

Parents responded to questionnaires on behalf of their child from birth, building an immediate picture of the child's circumstances. These questionnaires occurred at ages five, 10 and 16, continuing to detail many socio-economic and demographic factors about the child. The child sat multiple assessments at ages five, 10 and 16 measuring a range of cognitive abilities, including language, and their development into adolescence.

The same children were followed after leaving school, with measures recorded at ages 26, 30, 34, 38 and 42 regarding their educational attainment, occupational success, personal lives and much more.

Three benefits are immediately evident, relative to the Newcastle Thousand Families Study. The broader geographical sample increases the generalisability of findings, the larger sample size increases the statistical power for analysis and the later timepoint allows for more modern interpretation and measures, whilst still allowing for income measures well into adulthood.

The BCS70 benefits from the increased knowledge in cohort study design through the National Child Development Study (1958) managed by the same team, chronologically occurring between the Newcastle Thousand Families Study and the BCS70.

This data has previously been used for life course analysis of Language Impaired groups (Non-Specific Language Impairment and Specific Language Impairment, separated using cut-off scores on language and non-language tests at age 5), predicting outcomes such as poorer mental health and higher rates of unemployment at age 34 (Law et al., 2009; Schoon et al., 2010a). The primary methodology was logistic regression, comparing group odds ratios for dichotomised outcomes after controlling for a range of other factors.

There are four essential measures, or categories of measures, captured within this dataset which make it ideal for answering the research questions and complementing the previous analysis:

- A standardised measure of language, enabling comparability between participants on a continuous scale – The English Picture Vocabulary Test (Brimer and Dunn, 1962).
- Earnings in adulthood measured at multiple timepoints for the individual – Measures from ages 30, 34, 38 and 42.
- A wider range of available covariates – These are tested for relevance via a step-wise process to optimise the model specification.
- Mediator measures taken prior to the measures of income – Educational attainment is captured in adulthood and non-cognitive skills are captured in childhood.

Figures 3.1 and 3.2 overleaf details the attrition of the sample that has occurred over time within the study.

The measures within this attrition flow diagram are described according to the four categories of measures above, each described in turn within the following four sections. Comparisons are drawn between the full information sample used for analysis and the available data for each individual measure.

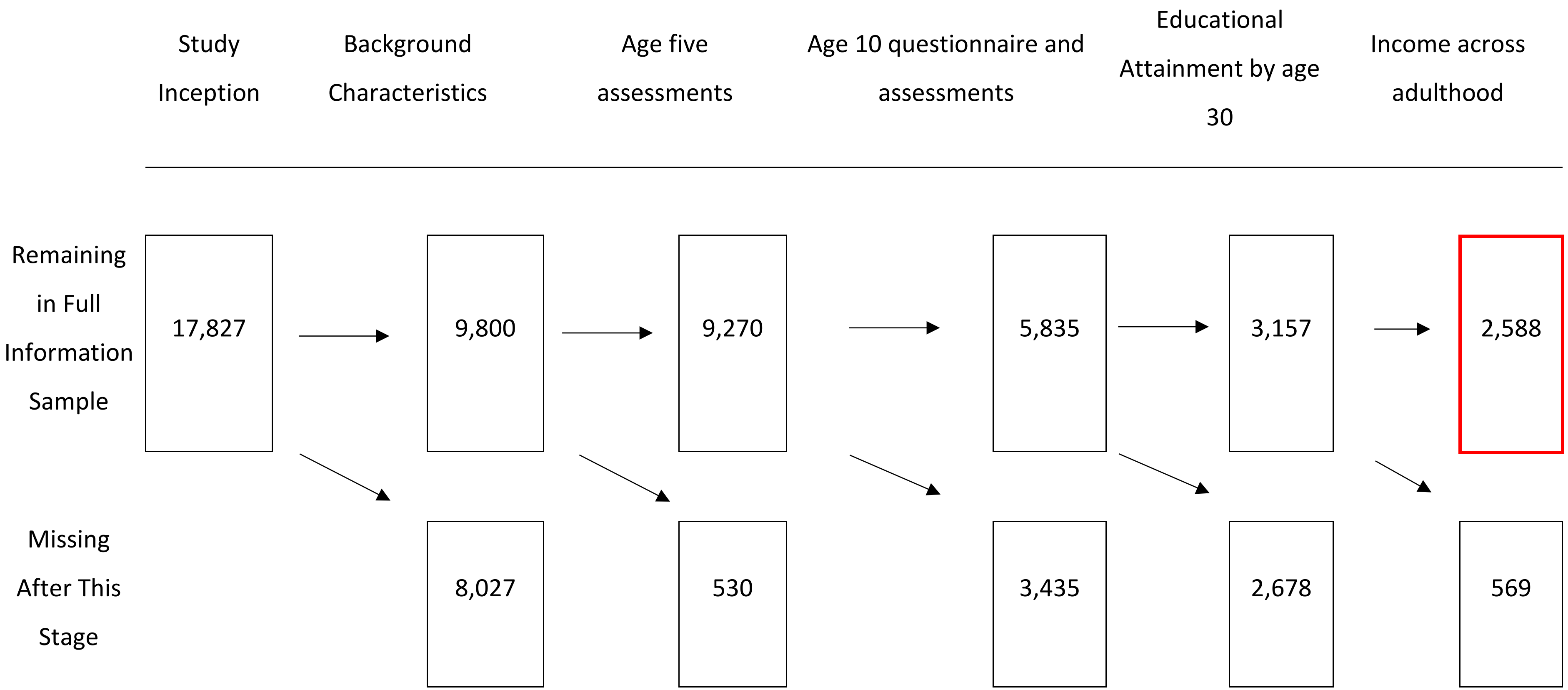


Figure 3-1 Sample attrition within the British Cohort Study

Note: Most missing data is missing from more than one variable. Therefore, exclusion of single variables has relatively little impact on sample size (see Figure 3.2). For example, of the 2,678 that drop out at the Educational Attainment timepoint, 1,305 are missing at all adult waves.

	Study Inception	Background Characteristics		Age five assessments		Age 10 questionnaire and assessments		Educational Attainment by age 30		Income across adulthood	
Data Missing Uniquely Across All Waves From		Father's Occupational Social Class	39	English Picture Vocabulary Test	125	Maths assessments	2	CASMIN Education Level	0		
		Gender	0			Reading Assessment	41			Income in adulthood	569
		Maternal Education	100	Copying Designs Test	3	Conscientiousness	352	Number of O-Levels	990		
		Paternal Education	198			Locus of Control	197				

Figure 3-2 Uniquely missing data by variable within the British Cohort Study 1970

Note: Number of O-Levels represents the most uniquely missing data, i.e. the removal of number of o-levels from the model would increase sample size by 990.

3.3.2 English Picture Vocabulary Test

Communication ability is measured in the early years using a cohort-specific modified version of the English Picture Vocabulary Test (EPVT) at age five. This measures receptive vocabulary (Brimer and Dunn, 1962), often referred to as 'comprehension', a crucial element of early language skills representing the child's ability to understand and process information delivered verbally (Law et al., 2017a; Taylor et al., 2013) and versions of this test have remained in regular use for decades (Roulstone et al., 2015). Receptive vocabulary is simply the number of words the child can understand and attribute meaning to, without necessarily needing to be able to produce those words themselves.

There is no available dataset with multiple measures of difference facets of language which also capture earnings at a sufficiently reliable stage of adulthood. The necessity of a large sample from this era brings limitations, the most important of which is the single limited measure of language. However, receptive vocabulary at age 5 is relatively preferable to other measures of early speech or language. Firstly, receptive vocabulary is highly cross-sectionally correlated with broader language abilities (Dionne et al., 2003). As we only have one early years language measure, receptive vocabulary is interpreted as representing broader language abilities, although care is taken to regularly refer back to the measure itself throughout this and the following analysis chapters.

However, this single measure does not allow for the modelling of differences between persistent and resolving difficulties. As discussed in the literature review (Section 1.2.4), receptive difficulties at age five are likely to be persistent (McKean et al., 2017a; Beitchman et al., 2008). Receptive difficulties are shown to be relatively stable from school age, with early severity also predicting later severity of difficulties (Law et al., 2008), as measured here on a continuous scale. The intention in the following chapters is to identify differences in early vocabulary, not to accurately identify SLCN – although SLCN could be one reason that a child does not perform well on the test.

Within the English Picture Vocabulary Test, the child is tested on their knowledge of word meanings. A word is read aloud and the child is shown a "plate" of four pictures. The child scores one point if they point to the correct picture. The child is shown a series of 56 "plates", each plate containing four pictures, and their score ranges from 0-56. The plates are arranged in order of increasing complexity.

To ensure that the child is responding to the test appropriately, scoring commences only once the child has answered five questions correctly. If this does not occur, a score of zero is recorded for that child. This can be seen in the distribution within Figure 3.3. After six incorrect answers are given within a run of eight plates, the test is ended. This reduces the amount of randomness as the child sits fewer multiple choice questions that are pitched above their receptive vocabulary level that may have led to the child guessing at random.

Scores are therefore distributed as either a 0 or any integer from 5-56. The distribution of scores is shown in Figure 3.3, with the strict upper ceiling of 56 questions explaining the short right hand tail. Figure 3.3 also shows how the distribution of the initial sample compares to the sample remaining with full information in our final analyses.

Children whose first and/or home spoken language was not English were removed from the analysis as per the instructions in the original test manual and the modern manual. (Brimer and Dunn, 1962; Dunn and Dunn, 2009)

3.3.2.1 Terminology

Whilst a receptive vocabulary assessment will not cleanly identify all the children with communication difficulties, it is an important feature in functional communication and is likely to relate strongly to issues experienced within school, conversation and the workplace (Bishop et al., 2016). Furthermore, many of the skills which underpin language or communication are highly correlated within high-functioning adults with a few dissociated exceptions in specific conditions. (Johnson et al., 1999)

A low score on the distribution shown in Figure 3.3 is considered to be a sufficient, but not necessary, condition for the presence of a language disorder. Children who cannot demonstrate receptive vocabulary on this test are unlikely to possess strong language skills in general. As the full distribution of scores is used, due to the limited diagnostic capacity of this sort of assessment, the continuous measure will be referred to as 'receptive vocabulary' or 'language' ability throughout this analysis but does not capture all facets of language ability.

This limitation is common in second-hand cohort data where quick, simple measures are used to maximise response rate within a study not specifically built to study language. This test has previously been used in this manner (Blanden, 2006) and the focus on receptive vocabulary is ideal due to its correlation with other facets of language (Dionne et al., 2003).

Table 3-1 Summary statistics of English Picture Vocabulary Test (EPVT)

Sample	N	Mean	Std. Dev.	Min	Max
All recorded	11,573	35.47	10.60	0	56
Complete Case	2588	37.60	9.86	0	56

Note: T-test reveals a significant difference in scores between the two samples

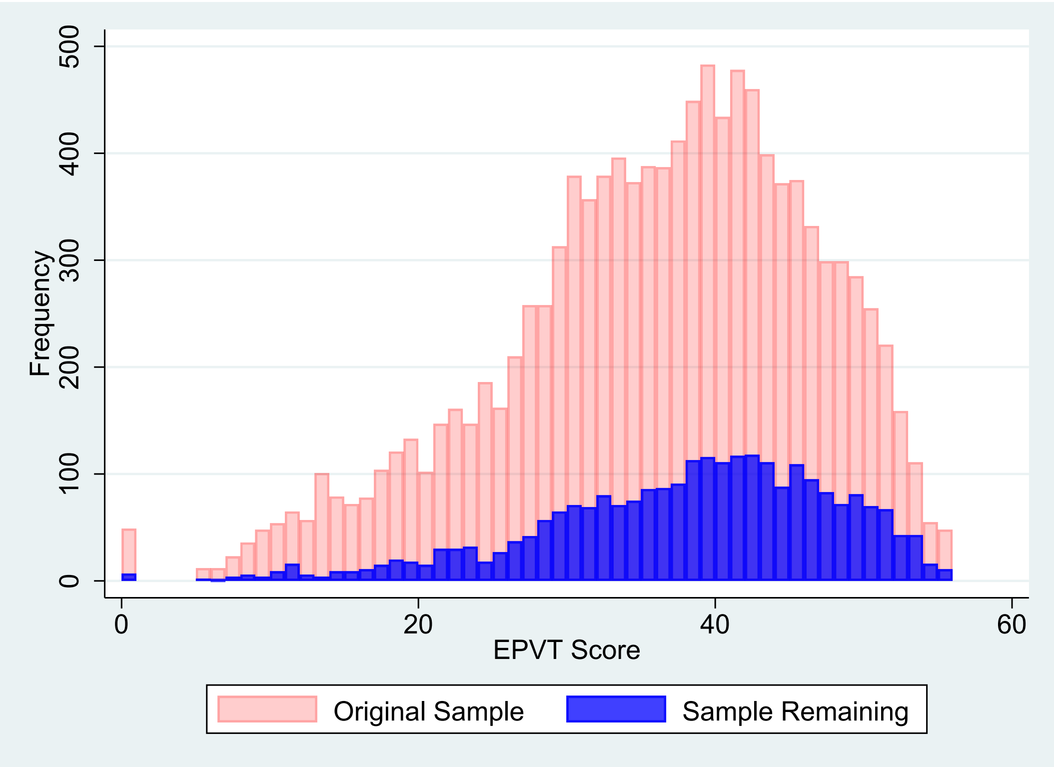


Figure 3-3 Histogram of EPVT score

Note: Sample is separated into those who sat the original EPVT and were not excluded from the analysis on the basis of the EPVT test, and those who remain in our complete case analysis.

The lack of scores from 1-4 is due to the specific testing procedure explained within the above section, with no score attributed to those who didn't get five consecutive words correct. There is a slight ceiling effect due to the limit of 56 questions but there is an otherwise normal distribution of scores.

3.3.3 Income

Our dependent variable is income in adulthood. Following the standard Becker-Mincer models of income in economics (Becker, 1964; Mincer, 1962), the income is logged. This accounts for the potential large non-linear variability at the highest end of the income distribution. By logging the income measures, we enable linear regression. As the income level is logged, this informs our interpretation of Figure 3.4, which shows the distribution of Time-Averaged Logged Income, and our subsequent results. A move of 0.01 along the scale represents a 1% increase in income. By extension, a move of 1 along the x axis represents the income level doubling.

Income was recorded in bands for every individual remaining in the study at each sweep in adulthood. The midpoint of the recorded band was used as an estimate of income at ages 30, 34, 38 and 42.

For this analysis, one outcome measure is created from these measures using time-averaging. The average of income at every recorded point is used to maximise accuracy by minimising the effect of temporary fluctuations in income (for example someone measured whilst temporarily unemployed if only one time point was used).

A secondary benefit to this measure is that we can maximise our sample size by averaging only the available data for each participant. If a participant only responds to one timepoint, this value is used in the analysis. This means that participants who answer all four timepoints may have a more accurate measure used than those who do not consistently respond to the survey questionnaires. This was preferred to a further reduction in sample size.

Income at 30 is considered to be the earliest reliable point for comparison as an indicator of maximum earnings (Gregg et al., 2014; Grawe, 2006) as, before this age, many people will be willingly working low-paid jobs or investing in their human capital via full-time education. For this reason, the measure of income at age 26 is not used within this analysis.

We use the time-averaged measure of income to represent earnings across adulthood. As there is variation in the reliability of earnings estimates in early adulthood by SES (Gregg et al., 2014), differences in the relationship between vocabulary and adult earnings over time are tested as a robustness check. This is to ensure that early vocabulary's relationship with earnings at 30 is not significantly different to the relationship with earnings by 42.

The distribution of time-averaged income midpoint estimation is shown in Figure 3.4 alongside the distribution of this within the full-information sample.

Table 3-2 Summary statistics of logged time-averaged income

Sample	N	Mean	Std. Dev.	Min	Max
All recorded	7275	9.63	0.69	4.88	12.65
Complete Case	2588	9.64	0.68	5.26	12.28

Note: T-test reveals no significant difference in scores between the two samples

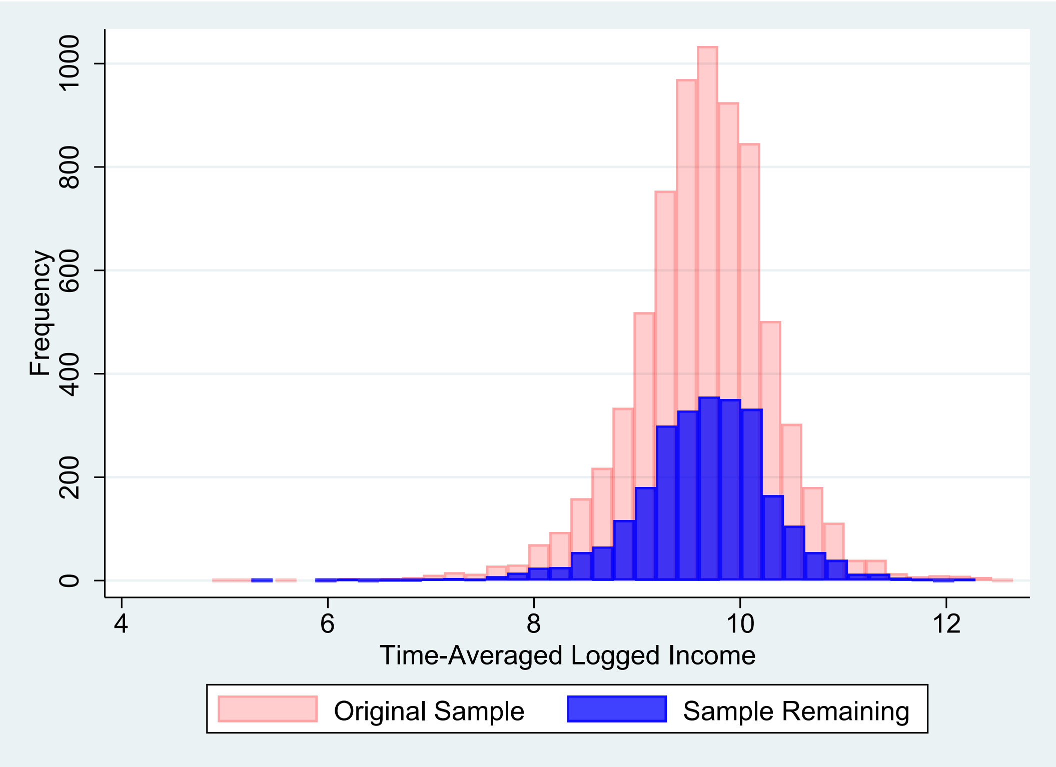


Figure 3-4 Histogram of logged time-averaged income by sample

Note: A score of 10 (X-axis) would represent approximately £22k annual income. A 0.01 movement along the axis would represent a 1% increase or decrease in this income (to approximately £22.2k or £21.8k).

3.3.4 Control Variables

For the purposes of this analysis, control variables are those which significantly predict the dependent variable but are not our specific mediation pathways (education outcomes and non-cognitive skills).

There are two primary types of control variable within the current analysis measuring the effect of early vocabulary on income. Firstly, factors relating to the background of the individual. Primarily, these will be correlated to SES and include factors such as the parent's level of education and occupational status. Many socio-demographic characteristics are present in the BCS70 as reported in Table 3.3. Secondly, factors relating to the child's non-communication capabilities which are likely to correlate with their language development and subsequent income. Many cognitive assessments were completed by the individuals at various ages. These are also listed in Table 3.3.

The BCS70 therefore contains a wide array of socio-demographic information through childhood which may predict our independent, mediating and dependent variables. This allows for a broader predictive model of income beginning in earlier life than the traditional Becker-Mincer model. However, we do not need to include Age or Age^2 in our analyses due to the birth cohort format of the data. We also exclude sectoral analysis of the income distribution due to limited data and within-individual variation over the time period of income analysed.

Due to incomplete data, it was impossible to control for every variable which could feasibly relate to early language development and later life income. Furthermore, inclusion of unnecessary controls reduces the accuracy of the coefficients attributed to the parameters we are interested in, the direct and indirect estimates of the effect of receptive vocabulary on earnings.

Therefore, a pragmatic statistical approach was taken. Twenty potential control variables (see Table 3.3 for a full list) were identified from the available data in childhood. Variables which related to the family background, the home environment or the child's cognitive ability were all contained within these twenty variables. With twenty potential control variables, a backwards stepwise regression process was employed for selection of the most important control variables. This is necessary to maximise sample size, and the backwards stepwise process was preferred to an All Possible Regressions Procedure due to the large number of potential combinations.

Variables which significantly predicted earnings were included in the final model. The backwards stepwise selection process is described in more detail in the following model selection section.

3.3.5 Model Selection

3.3.5.1 Backwards Stepwise Regression Selection

The English Picture Vocabulary Test was regressed as a predictor alongside every one of the twenty controls listed in Table 2. The least significant (highest p value, lowest t-statistic) variables are removed from the regression. This process continues until every control variable remaining has a p-value below a selected cut-off point.

A cut-off p-value of 0.107 was used, based upon the following calculation recommended in VanderWeele (2015):

$$\text{t-statistic} = \sqrt{2\log(K)} = 1.613 \text{ (3dp)}$$

Converted p value = 0.107 (2765 degrees of freedom)

K = Number of predictors = 20

This calculation balances the high probability that a model with many parameters (control variables tested as predictors of earnings) will find significant results through Type I (false positive) error. Therefore, the number of parameters informs the cut-off value.

This value, close to a 10% significance level, allows for the retention of seven control variables from the original model. This is a balance between an overfitted model, with twenty control variables and a resulting small sample size, or an underfitted model, with no control variables and a risk of omitted variable bias (which would be likely to cause an overestimate of the effect of vocabulary on earnings).

The results of this backwards stepwise regression selection process are shown in Table 3.3. Six control variables were confirmed to be predictive of earnings within the stepwise model.

Table 3-3 List of potential control variable and step-wise selection result

Note: P-values reported from backwards stepwise regression of control variables on earnings with p-values of below 0.107 retained for the mediation analysis.

Type of Control	***Variable***	***p value***	***Retained?***
Parental	**Father's Social Class 1970** **(6 Registrar General Levels + Unemployed)**	**0.053**	**Yes**
	Mother's Education level 1970 **(3 CASMIN levels)**	**0.021**	**Yes**
	Father's Education Level 1970 **(3 CASMIN levels)**	**0.000**	**Yes**
	Highest Parental Interest Expressed in Child's Education (Uninterested / Very Little Interest / Moderately Interested / Very Interested)	0.281	No
	Mother's Mental Health 1980 (Rutter Malaise, range: 1-2154)	0.430	No
Childhood factors at Birth	**Gender** **(Dummy variable for females)**	**0.000**	**Yes**
	Low Birthweight (<2500g)	0.326	No
	Short gestation period (<36 weeks)	0.344	No
Childhood factors aged 5	Has seen an SLT (Yes, currently or previously vs No, never)	0.738	No

	Copying Designs Test Score (Scale 0-8)	**0.025**	**Yes**
	Profile Test Score (Scale 0-16)	0.567	No
	Schonell Reading Ability level (1-4 Categorical)	0.382	No
	How often parents read to child weekly (0 – 7 days, self-reported)	0.385	No
Childhood factors aged 10	Child has seen an SLT (Yes, currently or previously vs No, never)	0.991	No
	Behaviour Problems (19-item Rutter Behaviour Scale 0-1854)	0.610	No
	Temper Problems (Scale 0-100)	0.759	No
	Friendly Maths Test score (Scale 1-72)	**0.000**	**Yes**
	Edinburgh Reading Test score (Standardised Scale, -3.23:1.95)	0.142	**Yes (after further analysis)**
	Child spoke to parents for over 5 minutes per day (Yes or No, self-reported)	0.397	No

Summary Statistics for all controls are available in the Appendix Table A3.1. CASMIN (Comparative Analysis of Social Mobility in Industrial Nations) levels are categorical variable classifications of education levels explained in detail (Brauns et al., 2003) in the Mediating Variables section.

3.3.5.2 Extension to the Backwards Stepwise Regression Selection

As the stepwise model contained many predictors which were inter-correlated, this increases the risk that important control variables may be excluded by a strict application of the Bonferroni point. Excluded variables which had substantial theoretical grounding for inclusion were re-tested in a reduced model to complement the original statistical procedure.

Due to their potential impact on education, parental interest in education, child behaviour and reading were further analysed for re-inclusion.

Parental interest in the child's education was expected to potentially predict income and could have been important for the pathways through early vocabulary and later education. However, no alternative specifications of the variable (individual parents' interests or the sum of the two parents' interests) were any more significant and would have reduced the available full information sample. This variable was therefore removed from the analysis.

Behaviour problems are commonly associated with language difficulties. Furthermore, it impacts on an individual's performance in school and is strongly related to socio-emotional regulation (Benner, 2002; Law and Elliott, 2009). This suggests it may play a relevant role in explaining pathways through both of our types of mediator. However, the Rutter behaviour scale, as reported by the cohort member's mother, did not significantly predict income in later life in the reduced model (i.e. after removing the other control variables excluded by the stepwise process) regardless of the inclusion or exclusion of language, education or non-cognitive skills. Therefore, the measure itself was deemed to be flawed for predicting earnings within this sample and modelling approach and was therefore removed from the analysis.

Finally, there were three variables included in the model relating to reading. These were tested in the reduced model (removing other non-significant controls) for joint and individual significance.

Parental reading to the child is commonly cited as a positive behaviour for parents to engage in with their children and one representation of a positive home learning environment (Bradley, 2002; Sénéchal, M. and LeFevre, 2001). However, whether modelled continuously (0-7) or trichotomised into 'Never', 'Sometimes' or 'Always' using the same data, the p value never reduced to below the Bonferroni cut-off. Given the presence of other parental characteristics, measures of vocabulary at 5 and reading ability at 10 subsequently included, it was considered safe to omit this variable.

Reading ability was measured in different ways at age 5 and 10, using the Schonell ability levels (Schonell, 1942) and the Edinburgh Shortened Reading Test (Godfrey Thompson Unit, 1978) respectively. The latter measure has a greater variation which may influence the stronger observed effect on earnings. This test uses word recognition as a measure of vocabulary, syntax, sequencing, comprehension and retention (Parsons, 2014).

In the reduced model, reading ability at age 10 was significant at the 1% level (p=0.010) and only led to a reduction of 41 participants (See Figure 3.2 for the attrition rates of all included variables), less than 2% of the remaining sample. The relationship between reading ability and income was consistent across randomly drawn sub-samples, suggesting that the significant effect was not caused by outliers not included in the backwards stepwise sample (due to missing data within the now-excluded variables).

The 1% significance level in the reduced model, the importance of reading ability to the literature relating early communication to later life outcomes (see Snow, 2020) and the minimal impact on sample size were all factors that cumulatively justify the inclusion of reading ability at age 10 in the mediational analysis model.

In total, seven control variables are established in a pragmatic manner, combining a statistical and theoretical approach. The parental occupation and education levels are treated as categorical or factor variables in each regression. Gender operates as a dummy variable for being female. The other selected variables were either count data or pre-determined scales and treated continuously as such.

These control variables each significantly predict the outcome variable and are included in each regression stage of the mediation analysis.

3.3.6 Mediating Variables

The Baron and Kenny (1986) definition of a mediator follows two criteria:

- The mediator must be predicted by the independent variable.
- The mediator must predict the outcome variable.

We are testing the mediation effect of two potential pathways, educational attainment and non-cognitive skills. The following sections cover the available measures of both mediator categories.

3.3.7 Education Attainment Levels as Mediators

Educational attainment is typically measured in the Becker-Mincer modelling of earnings through formal academic qualifications. The first available measure of income used is at age 30, and so we consider only the academic attainment achieved by this age in our analysis of time-averaged income. Particularly within this era, this is likely to be an accurate portrayal of the overall lifetime academic attainment of the vast majority of the sample.

Firstly, we measure the highest education level achieved by the individual. As the analyses require linear estimation, two binary variables are created to indicate two important milestones for educational achievement using the CASMIN (Comparative Analysis of Social Mobility in Industrial Nations) classification of education levels (Brauns et al., 2003). The CASMIN ratings separate our sample into three distinct categories from the original eight recorded in the survey, as shown in Table 3.4:

1. No Qualifications / D-E Grade CSEs or O-Levels / < 5 CSEs or O-levels
2. ≥ 5 CSEs or O-Levels / A-Levels / Higher Education Diploma
3. Degree / Post-Graduate Degree

This CASMIN method, first formatted in the 1970s, is widely used for modern international comparisons of educational systems (Kerckhoff et al., 2002). The CASMIN rankings are designed for ease of comparison across different educational systems, as is useful when modelling educational achievement within the 1980s for modern interpretation. Whilst more recent updates have needed to account for a wider range of vocational qualifications in advanced nations, this simple stratification is suited well for understandable interpretations of education qualifications achieved in the 1980s relative to modern schooling. (Brauns et al., 2003)

The above three-level factor variable is represented using two dummy variables. One for achieving a minimum of CASMIN Level 2, and a second dummy variable for achieving CASMIN

Level 3, with the baseline reference group being those in Level 1, the least academically qualified (=< 5 CSE).

Conditional expectations are modelled in the linear regression of these binary variables, with the outcome predicted by the regression representing the probability of an individual reaching that CASMIN level. The two levels allow for a differential mediating effect through achievement of a degree relative to a lower level qualification such as O-Levels or A-Levels. Increased vocabulary may be mediated through the increased marginal likelihood of belonging to the higher academic qualification group, with a separate portion of the mediation attributable to the odds of each categorical rise.

The distribution of CASMIN Levels within the complete case sample are shown in Table 3.4. Distributions of language scores by CASMIN levels are shown in Figure 3.5.

Given the literature regarding language development, primarily relating to adolescence and the completion of compulsory education, a further measure was included to capture the number of qualifications (O-Levels generally, but CSEs also included) obtained by the cohort member at age 15/16. This is a count variable ranging from 0 to 14. The distribution is shown in Table 3.5 with over 30% of the sample obtaining no formal qualifications and fewer than 10% of the sample obtaining 10 or more.

The fact that not everyone in the sample will have been made to sit formal examinations is important in the interpretation of these distributions. An unknown but potentially substantial percentage of those who obtain no formal qualifications will not necessarily have failed any exams. This was also the case within the Newcastle Thousand Families Study analysis in the previous chapter. In the modern educational context, you would not expect to observe a similar distribution and the number of qualifications is more likely to reflect the percentage of exams passed.

It is also worth remembering that we do not quantify anything detailed about each qualification, either in terms of grade or subject. Each pass is valued equally within our data and model. Modern analyses suggest that the financial returns to a degree vary significantly by subject, grade and university (Crawford et al., 2016), yet this level of detail is not observed in the data presented. Therefore, we expect considerable heterogeneity between individuals within the same CASMIN level or with the same number of O-levels.

Table 3-4 Distribution of highest academic qualification and CASMIN levels in complete case sample

Rank	Highest Academic Qualification Level Achieved by Age 30	N	% of Sample	N (Male)	N (Female)
0	None	784	30.3	434	350
1	D-E Grade CSEs	19	0.7	11	8
2	2-4 CSEs or O-Levels	0	0	0	0
3	≥5 CSEs or O-Levels	818	31.6	406	412
4	1 A-Level	51	2.0	26	25
5	2+ A-Levels	140	5.4	65	75
6	Higher Education Diploma	112	4.3	47	65
7	Degree	562	21.7	259	303
8	Post-Graduate Degree	102	3.9	58	44
Total		2588	100.0	1306	1282

Rank	Highest Academic Qualification Level Achieved by Age 30 (CASMIN 1-3 Level)	N	% of Sample	N (Male)	N (Female)
1	< 5 CSEs or O-Levels	803	31.0	445	358
2	O-Levels or A-Levels	1121	43.3	544	577
3	Degree +	664	25.7	317	347
Total		2588	100.0	1306	1282

Table 3-5 Distribution of number of o-levels (0-14) within complete case sample (two parts)

Number of O-Levels	0	1	2	3	4	5	6	7
N	820	154	146	118	108	144	140	195
% of sample	31.7	6.0	5.6	4.6	4.2	5.6	5.4	7.5
N (Male)	452	76	65	59	62	67	76	88
N (Female)	368	78	81	59	46	77	64	107

Number of O-Levels	8	9	10	11	12	13	14	15
N	283	244	146	59	22	6	3	0
% of sample	10.9	9.4	5.6	2.3	0.9	0.2	0.1	0.0
N (Male)	130	114	68	34	10	4	1	0
N (Female)	153	130	78	25	12	2	2	0

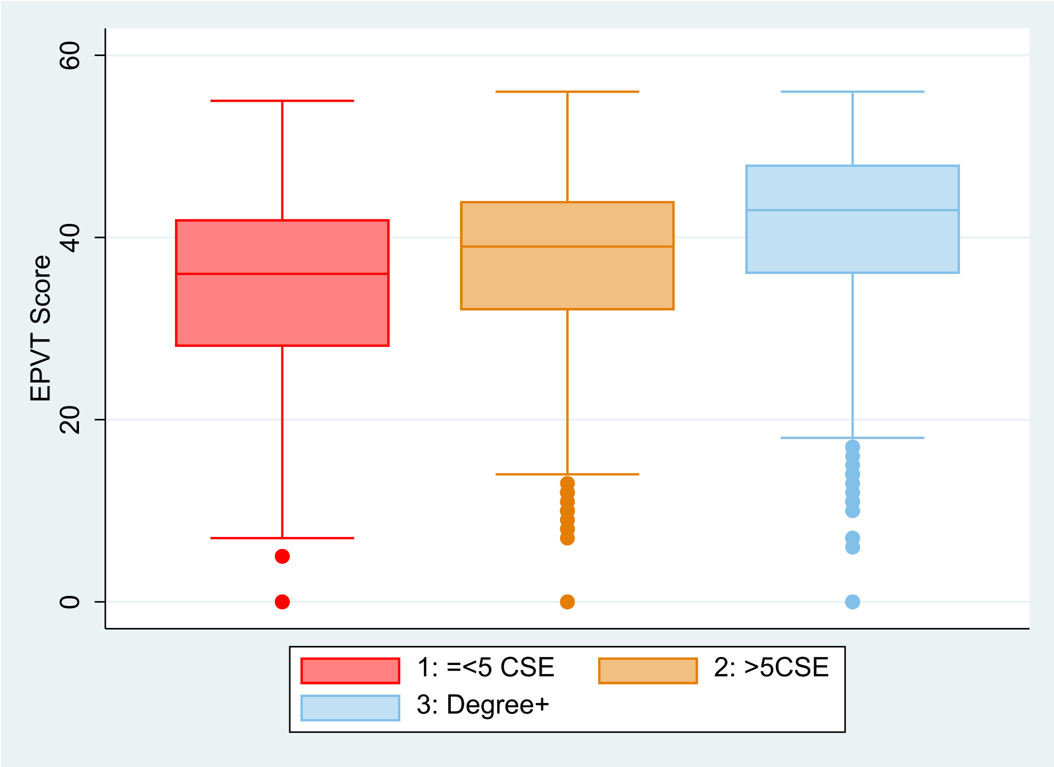


Figure 3-5 Box and whisker plot of receptive vocabulary scores by participant's highest academic qualification (CASMIN level) witihin complete case sample

Figure 3.5 represents receptive vocabulary scores at age five on the y axis, ranging from 0-56 with a mean of 37. These scores are differentiated by CASMIN level. This shows the vast individual differences in reality and the complexity of these associations. Using a single observation at age 5 to measure language ability with the potential for subsequent improvement or deterioration is potentially one part of the variation. Higher receptive vocabulary at 5 predicts a greater likelihood of membership of the more educated groups in later life, although individual variation is large.

There is demonstrably no ceiling to the academic achievement of a child with initially low or zero scores on the receptive vocabulary measure, as all three categories shown in Figures 3.5 include children with low y axis scores, although improving the vocabulary of these children would correlate with greater odds of attaining a higher level of education.

A similar relationship regarding the average effects of receptive vocabulary is demonstrated for O-level scores achieved at age 16 in Figure 3.6.

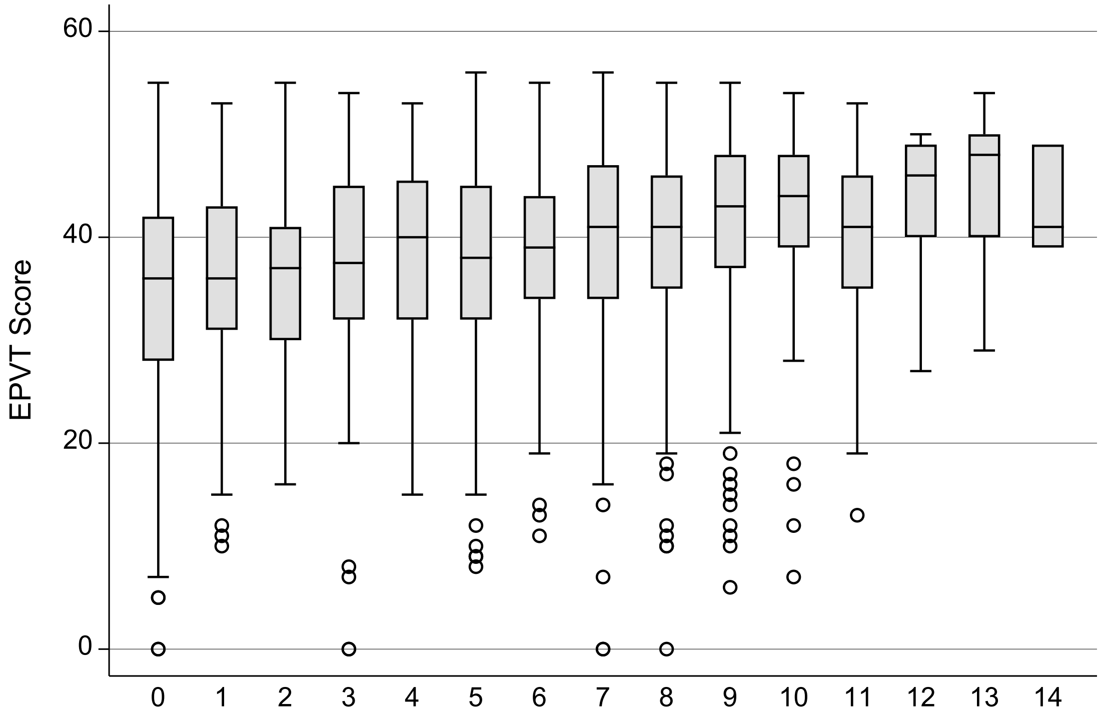


Figure 3-6 Distribution of receptive vocabulary (y axis) by number of o-levels (x axis) within complete case sample

As we would expect, a higher initial receptive vocabulary at age 5 shows a positive correlation with educational performance at age 16 through a general upwards trend shown in Figure 3.6. Again, there is a great deal of variation, although it is rare for a low language individual to obtain 11 or more O-levels or CSEs.

As expected from Table 3.5, there is a noisier trend (more randomness making the general trend less visible) in the highest groups with very few participants as it was rare to take (and pass) more than 12 examinations. The number of O-levels obtained is modelled continuously in the mediation analysis, predicting whether higher vocabulary scores predict higher earnings through an increased average number of O-levels achieved.

The mediation model captures whether this association between receptive vocabulary and education level predict later differences in average incomes across the distribution of vocabulary scores within our sample, although a significant result is not intended to be inferred as a reliable prediction of any individual's future.

3.3.8 Non-Cognitive Skills as Mediators

The second pathway tested for potential indirect effects of receptive vocabulary on earnings is through non-cognitive skills. The literature reviewed in Chapter 1 summarises a significant body of literature establishing the causal impact of non-cognitive skills on earnings in adulthood. Within the Chapter 2 analysis, we found significant effects connecting early speech and language development to the OCEAN non-cognitive skills in adulthood.

Locus of Control, the extent to which an individual feels in control of their outcomes, was not conceptualised within the Newcastle Thousand Families Study, but is captured at age 10 and 16 within the British Cohort Study 1970. The response rate in 1986, when the cohort members were sixteen years old, was handicapped by a teachers' strike (Heywood et al., 2015) meaning the non-cognitive skills used in this thesis were measured at age 10, before their full formation had completed in order to preserve the sample size of the overall model.

The individual's Locus of Control is measured via a 20-item binary response scale including 5 distractor questions, leaving a 0-15 scale. A higher score represents a more internalised Locus of Control meaning that the individual feels more in control of their own outcomes and less influenced by external circumstances. The "CARALOC" questionnaire used is available from the CLS website's Sweep Age 10 under Additional Documentation. The summary statistics are presented in Table 3.6 showing the full range of possible results were captured within our full information sample. The range of scores were normally distributed about the centre suggesting a relatively equal split of internalised and externalised individuals.

Within the British Cohort Study 1970, the OCEAN non-cognitive skills were not explicitly tested in childhood they were not conceptualised until McCrae and Costa (1987) and the common parameterisations in questionnaire form seen today were not available. Retrospectively, however, there have been attempts made to re-purpose questions which were answered which relate specifically to 4 of the Big 5 non-cognitive skills, namely omitting Openness to Experience.

Conscientiousness, Extraversion, Agreeableness and Emotional Stability have been retrospectively coded using items from a maternal questionnaire within the age 10 sweep of the BCS in 1980 (see Wehner et al., 2016 and Prevoo and ter Weel, 2012).

Table 3-6 Summary statistics for non-cognitive skills within the complete case sample

	N	Mean	Std. Dev	Min	Max
Locus of Control	2588	7.99	2.77	0	15
Conscientiousness	2588	305.18	79.58	7	397

However, of these 4 traits, only Conscientiousness was found to have a significant association with Income in our model (p = 0.050). Therefore, Extraversion, Agreeableness and Emotional Stability were removed from the model analysis.

Conscientiousness was measured using four items, each relating to one of the five facets (no item present for ‘industriousness’) of Conscientiousness as reported in Roberts et al. (2004), as well as Jackson et al. (2009). These summary statistics are also reported in Table 3.6, with the full item-level summary statistics reported in Appendix Table A3.1, as in Prevoo and ter Weel (2012).

In the analysis within Chapter 2, Conscientiousness was not significantly associated with early speech and language development in the multi-variate analysis. Association with the independent variable is a necessary criterion for a mediator (Baron and Kenny, 1986). Therefore, this association was also tested prior to conducting the main analysis.

Conscientiousness was not significantly associated (p=0.711) after account for the seven control variables established in Section 3.3.5.

Therefore, Conscientiousness was not included as a non-cognitive mediator in the mediation analysis. As Conscientiousness predicted Income in adulthood significantly, it was retained in our model as a control variable.

The full mediation model is depicted in Figure 3.7.

3.4 Methodology

3.4.1 Mediation Model

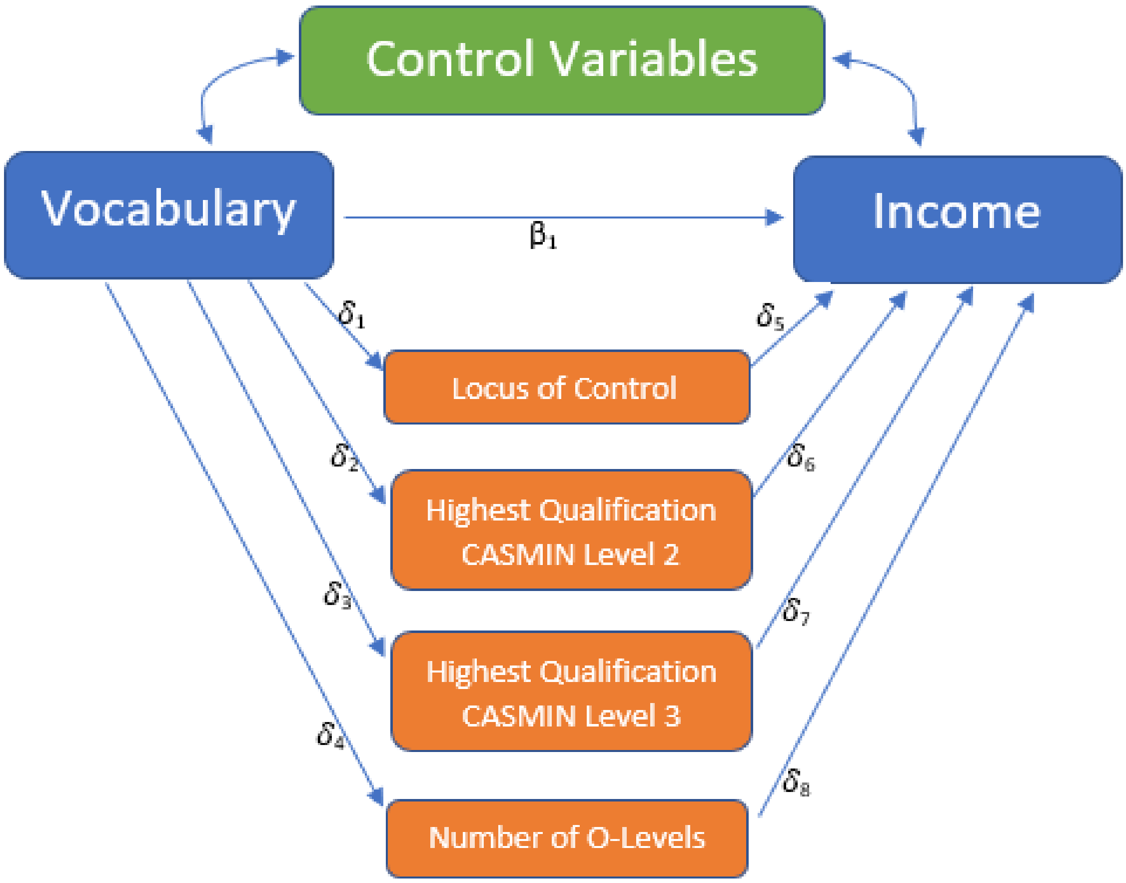


Figure 3-7 Mediation model path diagram with associated estimates

Note: Control variables = Gender, Maternal Education, Paternal Education, Father's Occupation, Copying Designs Test Score, Friendly Maths Test Score, Edinburgh Reading Test Score and Conscientiousness.

3.4.2 Structural Equation Modelling

The primary analysis will follow traditional Baron and Kenny (1986) mediation analysis. This is based on multiple Ordinary Least Squares regression estimates and is commonly known as the 'Product Method' or the 'Product of Coefficients Method'.

This is depicted in Figure 3.7, showing the pathways examined by this analysis.

The first stage regression predicts the mediator using the independent variable (language), and the list of eight control variables. Ordinary Least Squares linear regression is used (Ramsey

Reset and Pregibon Link tests of the functional form were used to confirm this approach was appropriate). This stage is conducted separately for each mediator using the same fixed, full information sample.

The second stage predicts the outcome using the independent variable, the mediator(s) and the control variables.

As we are testing multiple mediators, there will be multiple first stage regressions, one per mediator. They will all be included simultaneously in the regression on earnings to avoid double counting through correlations across mediators.

This is a total of five linear models, the fifth model capturing five estimates from Figure 3.7. Stata's Generalised Structural Equation Modelling programme was used to capture the paths depicted in Figure 3.7. Full information maximum likelihood estimation is used within this programme to generate the linear estimated effects of each regression and to control for the correlation of error terms across equations. Full information in this case means only the complete case sample is used on all five regressions (such that the maximum likelihood estimation is equivalent to a seemingly unrelated regression technique).

The final model specification allows for a direct correlation between vocabulary and earnings to be estimated. In combination with the four stage one regressions, we also obtain estimates of the indirect correlations through each of our mediators.

The direct correlations are measured using the B, or estimated coefficient associated with the independent variable in the final regression. The indirect correlations are measured by multiplying the δ, or estimated coefficient, associated with the independent variable in the relevant mediating regression and the δ associated with the mediating variable in the final regression.

Although the mediators are scaled differently, this is neutralised by the product method. Locus of Control has a larger scale than CASMIN Level 2, for example. This results in relatively larger coefficients in the first stage estimation of Locus of Control and equivalently lower coefficients estimating the impact of Locus of Control in the second stage estimation of Income.

Any total effect of vocabulary on earnings is thereby decomposed into the direct effect and the indirect effects through non-cognitive skills and education.

3.4.3 Sample attrition

All regressions are conducted using one fixed, complete case sample, as summarised in the attrition flow chart in Figure 3.1 and used for all summary statistics. Due to the length of time over which we are using data on the individuals, only a small proportion of the sample have full data available. In particular, there are a large number of missing observations in the Number of O-Levels/CSEs obtained variable due to collection problems relating to strike action in the 1980s (Heywood et al., 2015), as displayed in Figure 3.2. There is further missing data where an individual responded to some but not all of the questionnaire at various time points.

For each variable, including the English Picture Vocabulary Test, performance is higher on average within the complete case sample. This impacts our interpretation as the remaining data is not nationally representative. This non-random attrition is typical for missing data in longitudinal datasets and modern cohorts have used over-sampling of particular populations to deal with this.

Three methods are proposed to test the robustness of the primary results derived from the complete case analysis.

Firstly, multiple imputations by chained equations is used to create imputed values for missing data. This is the least bias method for dealing with missing data (White and Carlin, 2010). The range of plausible values allow through multiple imputations is preferable to single imputation (Rubin, 1987).

Secondly, the number of O-levels achieved by the individual is removed from the analysis, as this is the most uniquely missing variable within the model. This variable is the most likely to be missing at random, given the teacher's strike impacting data collection in 1986 (Heywood et al., 2015).

Thirdly, the use of a complete case sample for all five regressions is known as 'listwise deletion'. An alternative method, 'pairwise deletion' is tested where the complete cases available for each regression are used as the analytic sample. For example, if only income and Number of O-levels were missing for a participant, the regressions predicting Locus of Control and the CASMIN levels would still contain that participant. This is the standard approach in Stata's Generalized Structural Equation Modelling programme, overruled initially in favour of the complete case analysis.

Each of these methods test the robustness of the initial analysis to changes in the sample. Further, they increase the sample size, allowing for higher power in the regressions and any changes in results are interpreted with this in mind.

3.4.4 Testing for significance

The direct effect can be tested for significance using the standard linear regression process (p value and t-statistic associated with the β coefficient), but more options exist for mediated pathways reliant on more than one regression.

In traditional Baron and Kenny (1986) mediation, overall indirect effect significance is reliant upon significance at each individual stage of the analysis, whereas the Sobel (1982) test can be more appropriately used to measure joint significance of an indirect pathway. However, Hayes (2009) recommend a bootstrapping approach as the most effective method of analysis and this will be the basis of our analysis. The Sobel (1982) and Baron and Kenny (1986) measures will be used as supplementary robustness checks.

The diagram in Figure 3.7 represents the logical model being measured. $\delta_1 * \delta_5$ would represent the indirect effect from vocabulary through an internalised Locus of Control onto earnings. The estimated indirect effect is drawn from the complete case sample. A bootstrapping approach is used to generate the confidence intervals (CI). Bootstrapping randomly samples with replacement, with many repetitions (10,000 used here). As the sample will randomly vary in each repetition, 10,000 estimates of the indirect effect ($\delta_1 * \delta_5$) are created and ordered by size. The lowest and highest 250 estimates (2.5% at the two tails) are removed from the distribution to create a 95% CI around the initial effect size estimated. If this confidence interval does not contain zero, then the effect is considered to be significant at the 5% level. Similarly, if the lowest 50 and highest 50 estimates were removed in the same manner, we could test significance at the 1% level. This distributional CI is created automatically in Stata 15 to create a p value for each indirect effect. Bootstrapping is also recommended as the preferred method for significance testing of the direct effect.

Appendix figures A3.1 to A3.4 demonstrate the correlation between early vocabulary and our mediating variables (Internalised Locus of Control and Educational Attainment) and outcome (Time-Averaged Income) but the heterogeneity is worth noting. We are modelling average effects across a continuous scale to measure whether receptive vocabulary tends to predict higher earnings and whether this is through a tendency towards internalised Locus of Control and higher educational attainment. It is clear from our data visualisation that some individuals with poor initial vocabulary scores earn well above the cohort average. Similarly, some low earners had initially high language scores. The same heterogeneity is evident in CASMIN Education levels, with each level containing individuals from across the language spectrum. This finding is echoed throughout the literature, that SLCN (of any form, including but not restricted to receptive language problems) can be predictive of poorer life outcomes whilst not being an unavoidable sentence. For many years, researchers have attempted to understand the mechanisms explaining this potential for resilience (Conti-Ramsden and Durkin, 2016).

3.4.5 Model Specification

$$M_{1-4} = \alpha_{1-4} + \delta_{1-4}EPVT_i + \gamma_{1-4,C}C_i + \omega_{1-4,i} \qquad (1) – (4)$$

$$Y_i = \alpha_5 + \beta_1 EPVT_i + \delta_{5-8}M_{1-4,i} + \gamma_C C_i + \varepsilon_i \qquad (5)$$

$$EPVT = Language\ Score\ at\ 5$$

$$M = Vector\ of\ Mediators$$

$$C = Vector\ of\ Control\ Variables$$

$$\omega = Stochastic\ Error\ terms\ from\ mediation\ models$$

$$Y = Time - Averaged\ Income\ Estimate$$

$$\varepsilon = Stochastic\ Error\ term\ from\ full\ model$$

In stage 1, the mediator (1-4 represent our four mediators) is predicted by a constant, the receptive vocabulary score at age five, and the 8 control variables (C). An error term is produced for every participant in each equation. The key parameters obtained are our δ parameters, the linear coefficient between language and the mediator.

In Stage 2, the outcome (time-averaged income) is predicted by a constant, the receptive vocabulary score at age five, the four mediator terms (1-4 representing the four mediators) and the eight control variables (C). An error term is produced for every participant. The key parameters obtained are our β parameter, estimating the linear coefficient between language and income, and our δ parameters, the linear coefficient between mediator and income.

The direct effect of language on lifelong income is estimated by β_1.

The indirect effect of language on income through a potentially more centralised Locus of Control in early childhood is estimated by δ_1* δ_5.

The indirect effect of language on income through a potentially higher maximum academic qualification level achieved by age 29 is estimated by δ_2* δ_6 for achieving CASMIN Level 2 and estimated by δ_3* δ_7 for achieving CASMIN Level 3.

The indirect effect of language on income through a potentially higher number of O-Levels or CSEs obtained at the end of compulsory educations is estimated by δ_4* δ_8.

The total indirect effect of language on income through education can therefore be summated as the cumulative indirect pathways through the attainment of CASMIN levels 2 and 3 as well as the number of O-levels, written mathematically as δ_2* δ_6 + δ_3* δ_{7} + δ_4* δ_8 or $\sum_{i=2}^{4} \delta_i * \delta_{i+4}$.

3.5 Research Questions and Hypotheses

1. Does a child's vocabulary predict their earnings in adulthood?

We hypothesise that higher childhood vocabulary will relate to higher later earnings ($\beta_1 > 0$).

2. Does performance in the education system mediate any relationship between early vocabulary and earnings?

We hypothesise that educational attainment may mediate some of the effect of vocabulary on later earnings ($\sum_{i=2}^{4} \delta_i * \delta_{i+4}$. > 0).

3. Are non-cognitive skills developed in later childhood mediators of any relationship between early vocabulary and earnings?

We hypothesise that non-cognitive skills, as measured by locus of control, in childhood may mediate some of the effect of vocabulary on later earnings (δ_1* δ_5. > 0).

3.6 Results

Table 3-7 Mediation analysis results

* = p < 0.1, ** = p < 0.05, *** = p < 0.01

	Direct Receptive Vocabulary	Indirect through Locus of Control	Indirect through CASMIN Level 2	Indirect through CASMIN Level 3	Indirect through Number of O-Levels	Total Indirect Through Education
Effect on Time-Averaged Income	0.003**	0.00005	0.00001	0.00009	0.0001	0.0003
	(0.001)	*(0.00006)*	*(0.00006)*	*(0.0002)*	*(0.0001)*	*(0.0003)*

Note: β *and Product* δ *Coefficients reported with bootstrapped standard errors in parentheses. Generalised Structural Equation Modelling conducted in Stata GSEM programme.*

3.6.1 Research Question 1

Does a child's vocabulary predict their earnings in adulthood?

Null Hypothesis H_0: $\beta_1 = 0$

Alternative Hypothesis H_A: $\beta_1 \neq 0$

We reject the null hypothesis at the 5% level and conclude that early receptive vocabulary has a significant direct effect on earnings in adulthood.

3.6.2 Research Question 2

Does performance in the education system mediate any relationship between early vocabulary and earnings?

Null Hypothesis H_0: $\sum_{i=2}^{4} \delta_i * \delta_{i+4} = 0$

Alternative Hypothesis H_A: $\sum_{i=2}^{4} \delta_i * \delta_{i+4} \neq 0$

We cannot reject the null hypothesis at the 5% level and conclude that there is no observed indirect effect of receptive vocabulary on earnings through the joint education pathways.

3.6.3 Research Question 3

Are non-cognitive skills developed in later childhood mediators of any relationship between early vocabulary and earnings?

Null Hypothesis H_0: $\delta_1 * \delta_5 = 0$

Alternative Hypothesis H_A: $\delta_1 * \delta_5 \neq 0$

We cannot reject the null hypothesis at the 5% level and conclude that there is no observed indirect effect of receptive vocabulary on earnings through an internalised Locus of Control in later childhood.

3.6.4 Effect Ratios

Approximately 90% of the total observed effect is direct (Direct: Indirect Ratio ≈ 9:1).

The statistically non-significant remaining 10% is primarily through Education (Indirect Education Pathway: Indirect Locus of Control Pathway Ratio ≈ 4:1).

3.6.5 Standardised Effect Sizes

By standardising the predictor and outcome variables, we find that a one standard deviation increase in early receptive vocabulary correlates with a 0.046 standard deviation increase in income (CI 0.008 – 0.085). This is characterised by Cohen (1988) as a small effect size, or a very small effect size by Sawilowsky (2009). However, any significant effect is a strong signal of the importance of communication ability given the enormous number of factors which will affect an individual's income in the 25 years between the receptive vocabulary measure and our first observation of income.

The approximate 0.046 standard deviation increase in income resulting from a 1 standard deviation change in language translates to approximately 3.2 percentage points of income (CI 0.53 – 5.91). A 3% increase in the 2019 median UK income of £29,600 (ONS 2020) would represent £888 additional pre-tax salary per year. Over the course of the 12 years we measure in adulthood, or potentially over a full career, this additional income could contribute greatly to an individual's quality of life, financial security and tax returns. Whilst the standardised mean effect size can be characterised as small by Cohen (1988), this is not an intervention study and these small effects are clearly stable over an extensive period of time.

The estimate associated with a standard deviation increase in receptive vocabulary (3.2 percentage points) may be more representative of the potential effects from increased interventions than the estimate associated with a unit increase in receptive vocabulary (0.30 percentage points). This is due to the estimated effect of an intervention being close to 1 mean standard deviation relative to control groups (delayed or no intervention) although there is huge variety in study findings around this, particularly relating to receptive vocabulary improvements (Law et al., 2004). To accurately estimate the potential benefits of an intervention using the analysis ran in this study would require far greater evidence of the potential to screen, identify and treat receptive vocabulary deficits as discussed in the literature review (see Section 1.3.1) in more detail.

3.6.6 Robustness Checks

In order to test whether the observed correlation was consistent over the 12 years with 4 observations of income, the model specification in Figure 3.7 was ran separately for logged income at ages 30, 34, 38 and 42. The estimates obtained at each separate time point are compared in Appendix Figure A3.5 with no significant differences between any of the given time points. This suggests that our correlations between communication and income are consistent between the ages of 30 and 42.

The results presented in Table 3.7 are estimated using bootstrapped confidence intervals as the most robust method available (Hayes, 2009). Bias Corrected and Acceleration Corrected Confidence Intervals (95%) were tested for each of the derived effects. Significance levels were consistent across all methods, demonstrating a direct effect but no significant indirect effects. This shows that the results were not impacted by bias (proportion of bootstrap estimates that were below the observed estimate) or acceleration (proportion of skewness in the distribution of the bootstrapped estimates). These results are reported in Appendix Table A3.2.

The results reported in Table 3.7 are also robust to the traditional Baron and Kenny (1986) confidence intervals within each pathway estimate, and to the Sobel (1982) method of joint significance of the two pathways combined. Appendix Tables A3.4 and A3.5 report these results.

Restricted sampling was attempted to test for differential effects between males and females. The findings in Table 3.7 were robust in either sample, although the limited sample and small effect sizes make it unlikely that a significantly different effect would be observed.

3.6.7 Missing data

The remaining full information sample within the BCS70 (n=2588) represents only a small portion of those with recorded income and vocabulary data (n=6718) suggesting that our mediating and control variables were significantly affecting our potential to measure the association between vocabulary and earnings.

A simple bi-variate regression would significantly over-estimate (4 to 5 times the effect size or β in the full controlled model specification) the actual effect of vocabulary on earnings by ignoring the relevance of our control variables. However, the reduction of the sample does not significantly change the β observed in a bi-variate regression. This is shown clearly in Appendix Figure A3.6.

Multiple potential solutions were attempted to test the effect of increasing sample size without reducing the complexity of the model and risking an over-estimation of the parameters of interest.

Firstly, multiple imputation by chained equations was attempted to impute some values for variables which were missing on specific individuals in order to include participants in our analysis where they had responded to the majority of the relevant survey questions.

Secondly, as number of O-levels had the lowest response rate of any variable, the model was tested without this pathway. CASMIN Level 2 may pick up some of the differential effect from removing this variable.

Thirdly, pairwise deletion was used for each individual regression, as opposed to an overall complete case analysis.

3.6.7.1 Multiple Imputations by Chained Equations

50 replication datasets were created containing potential imputed values for the non-monotone missing data (different variables missing from each individual with no consistent pattern). All variables within our mediation model were used to provide produce the imputations presented in Table 3.8. This increased our sample from the complete case sample of n=2,588 to an imputed sample of n=4,385.

The same model specification was estimated to predict our five regressions and our range of δ's and the direct effect β, producing the results reported in Table 3.9.

Table 3-8 Summary of Multiple Imputation by Chained Equations (50 replications)

Variable	Responded	Incomplete	Imputed	Total
Number of O-Levels	6205	12905	990	19110
Time-Averaged Logged Income	7275	11835	569	19110
Locus of Control	11838	7272	421	19110
Reading	9586	9524	109	19110
Conscientiousness	12646	6464	53	19110
Maths	11630	7477	3	19110

Table 3-9 Mediation results within imputed dataset

p<0.1 = *, p<0.05 = **, p<0.01 = ***

	Direct Receptive Vocabulary	Indirect through Locus of Control	Indirect through CASMIN Level 2	Indirect through CASMIN Level 3	Indirect through Number of O-Levels	Total Indirect Through Education
Effect on Time-Averaged Income	0.003***	0.00004	0.00003	0.0001	0.00015	0.0003
	(0.001)	*(0.00004)*	*(0.00004)*	*(0.0002)*	*(0.0001)*	*(0.0002)*
(Original Complete Case Estimates)	0.003**	0.00005	0.00001	0.00009	0.0001	0.0003
	(0.001)	*(0.00006)*	*(0.00006)*	*(0.0002)*	*(0.0001)*	*(0.0003)*

Note: Multiple Imputations command (mi estimate) incompatible with Bootstrap (bootmm). Sobel Test confidence intervals used instead, averaged across the 50 datasets using Rubin's Rule as explained at "https://stats.idre.ucla.edu/stata/faq/how-can-i-compute-indirect-effects-with-imputed-data-method-2/". *Original Table 3.7 results included for comparison.*

The above results are consistent with those predicted by the original bootstrapped model in the full information dataset. Whilst this does not confirm the results to be generalisable to the population as there is still a significant amount of missing data, it does make the initial results more reliable as there is an extra 60% of participants with sufficient responses to impute the remainder who can now be included in our model.

The confidence interval around the direct effect has narrowed ($p=0.004$) and the β increased from 0.0029 to 0.0032 suggesting slightly stronger evidence of a direct effect.

Similarly, the total indirect effect through Education is higher overall although it is still not statistically significant ($p=0.205$).

In summary, whilst there are some very slight differences in effect sizes and coefficients, the multiple imputations estimates serve primarily to confirm our initial findings in the complete case sample.

3.6.7.2 Number of O-Levels removal

As the main variable in need of imputation was the Number of O-Levels achieved, a specification was tested without this variable under the assumption that CASMIN Level 2 would capture some percentage of these effects. The same model specification is applied although $\delta_2 * \delta_6$ are no longer estimated due to the removal of this mediator. The results for this specification are presented in Appendix Table A3.3, showing consistency with the initial estimates.

3.6.7.3 Pairwise Deletion

Pairwise deletion allowed for complete case analysis within each individual regression of the five linear regressions. Research Questions 1 and 3 are unchanged by this method, although there is a slightly higher non-significant indirect effect through locus of control.

However, there is a significant indirect effect of vocabulary on earnings through education levels at the 5% level when using a pairwise deletion method. Figure 3.8 below shows the change in effect when utilising a pairwise deletion method. The actual change in effect size is relatively minor, although the confidence interval is drawn in enough to observe a significant effect. The direct effect remains significantly larger than the overall indirect effect.

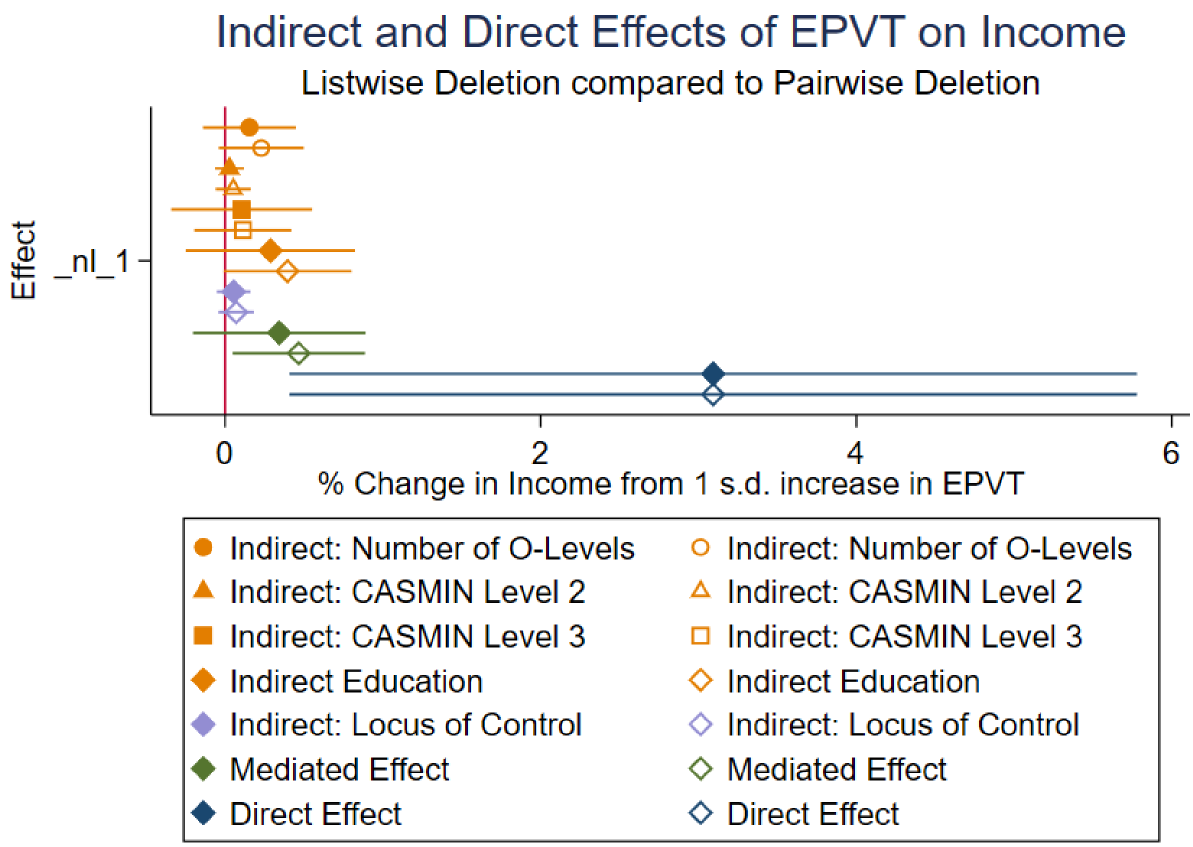


Figure 3-8 Listwise and Pairwise Deletion approaches to missing data

Note: Hollow point estimate markers denote pairwise deletion methods. The direct effect estimate is unchanged as all variables are included in the fifth regression, meaning the same complete case sample is used.

3.7 Discussion

The results presented here represent the first analysis to our knowledge of the effects of early language ability on earnings in adulthood. We find that a standard deviation increase in receptive vocabulary at five correlates approximately to a three percentage point increase in income per year. This effect is significant at the 5% level in all pooled sample specifications although the confidence interval tends to be wide, suggesting anything from a 1 to 8 percentage point increase in income may be the underlying true β.

This effect is estimated whilst controlling for gender (dummy variable), SES (father's and mother's education as well as father's occupation at birth), other tests of cognitive ability (copying designs score at 5 as well as maths and reading ability at 10) and non-cognitive ability (conscientiousness at 10) which all add to our prediction of income.

Furthermore, this effect is estimated whilst controlling for multiple potential pathways through which early vocabulary could have additional effects on earnings (internalised locus of control at 10 as well as the highest education level achieved and the number of O-levels obtained).

Whilst we cannot control for all factors which correlate with both early language ability and later life income, we do cover a wide range due to the depth of the data used. A bi-variate prediction of income using the same fixed sample with only our receptive vocabulary measure as a predictor estimates an effect size around 5 times as large showing that these controls are clearly an important part of the more general correlation. This is shown in Appendix Figure A.3.6 clearly. The huge impact of control variables reflects the context in which vocabulary is developed, and the importance of many proximal and distal factors in shaping an individual's life. Without these control variables, simple cross-sectional comparisons suggest much more stark differences in the life outcomes of children with different levels of early vocabulary.

Even with regular measurements and structural models, we would not fully capture an isolated effect perfectly in a child's life, as there are so many consequential interactions. The presence of these interactions and lagged effects on other variables motivated the use of a mediation model. However, the effect size of the early vocabulary measure on income in adulthood was small, albeit substantial over the life course. In statistical terms, this makes it unlikely that a mediation effect will be observed given that the mediators are not intrinsically linked to vocabulary (i.e., you wouldn't expect the majority of the impact to be via mediation). This is partially limitation of the measures used; firstly, the vocabulary measure is continuous and includes an element of multiple choice, meaning that the coefficient observed is likely to be small with wide confidence intervals; secondly, the education variable is binary, and so the impact of early vocabulary has to impact income specifically through a narrowly measured qualification level; thirdly, the income measure is categorical, and so the impact has to specifically move the individual into a new income bracket to be observed. The limits of these measures may reduce the likelihood of observing a statistically significant direct or indirect effect.

Given these limitations, the presence of a significant direct effect allows us to confidently conclude that receptive vocabulary, even measured as early as 5, is positively associated with later earnings in life although the effect size may be categorised as small. It is important to note that whilst income is important in life, it is not the same thing as quality of life. When measuring differences between Language Impaired and Typically Developing groups, there is evidence of differential satisfaction rates with the same education outcomes (Durkin et al., 2009). Many factors influence labour market decisions and a group with poorer functional communication skills may willingly self-select into jobs with fewer financial and greater non-pecuniary benefits from their perspective. Alternatively, having fewer overall job opportunities may mean that a group with poorer functional communication skills might be forced into jobs with fewer non-pecuniary benefits due to lack of choice and that this negative outcome is not captured by a measure of income. Our results only measure the financial aspect of the labour market and cannot explicitly measure job satisfaction.

The overall finding on our direct effect backs up our initial findings from the Newcastle Thousand Families Study data, suggesting that these effects are robust to a significant change in the measure of communication used, the measure of income used, the range of individual and family level controls used and the context of the data historically.

Our results do not suggest the presence of a significant indirect pathway through early non-cognitive skills measured at age 10. This may be explained by the poor retrospective measurement of non-cognitive skills from questionnaires not designed to elicit these traits in terms of the Big 5, although there is no effect found through internalised Locus of Control either. As the literature shows significant differences in self-efficacy between those with and without Language Impairment, more research is needed to establish the timings of these correlations and whether age 10 would be too early to detect these traits and their associations with language.

Furthermore, we find no significant mediation effect of education, measured in multiple ways. There is a direct effect between education and later life income (combined $\delta_{6\text{-}8}$), p=0.000), but no significant portion of this effect is explained by initial receptive vocabulary at age 5. Whilst educational attainment differs by language profile in both the literature and our data, the inclusion of multiple control variables significantly tempers this association.

The remaining sample used for these specifications represents only a small portion of the initial sample representing the population of the UK born in 1970. There is evidence that the attrition in the sample is non-random, reducing the inference we can draw from the results presented. Multiple imputations methods can account for some of the missing data without impacting the observed results. However, it is likely that the individuals for whom we can impute values to missing variables more closely resemble the individuals within the complete case sample than those for whom we have too little data to generate imputed values, as they must have responded to at least some of the adult questionnaires.

Whilst the remaining sample may not be representative of the UK population anymore, it is still a very large sample size relative to those generally found in the literature. This is necessary for obtaining population level estimates in the context of a 25-year gap between the observation of their vocabulary development and their first income measure where you would only anticipate small effect sizes.

3.8 Conclusions

This thesis set out to accomplish three broad aims in improving understanding of the life experiences of individuals with speech, language and communication needs. The three aims were as follows:

1) To utilise historical data to highlight the enduring impact of early of early speech and language development over the life course.
2) To expand the range of econometric methodologies and outcome measures used in the field to increase understanding of the breadth of these impacts.
3) To make appropriate, evidence-based recommendations for research and clinical practice.

This chapter confirmed the presence of an association between early receptive vocabulary and later life earnings. A one standard deviation increase in receptive vocabulary at five correlates to an approximate 3.2% increase in annual income. No significant mediating effects were observed through either educational attainment, or locus of control.

The British Cohort Study 1970 was employed for the current analysis due to the presence of a relatively stable language measure in childhood and a number of observations of income across adulthood. Generalized Structural Equation Modelling was used to fit a linear mediation model with a range of control variables, including factor variables.

Whilst the direct effect observed supports the literature surrounding the importance of early language development for economic outcomes, further work is needed to understand the mechanisms behind this association in order to make appropriate conclusions. Therefore, the following analysis employs a decomposition technique, commonly used to analyse wage differentials, the Oaxaca-Blinder Decomposition. The same cohort and data are used for consistency, with the intention of further understanding the reasons why children with poorer vocabulary at age five are likely to earn less in adulthood.

4 Earnings Differences Between Vocabulary Groups: are skills equally rewarded?

4.1 Introduction

Chapters 2 and 3 demonstrate that earnings tend to be higher on average for individual with stronger early communication skills. Previously, the analysis has only allowed for discussion of the theoretical impact of increasing early communication skills in isolation. However, in reality communication ability interacts with many of the other variables which predict earnings, as shown by Appendix Figure A.3.6 – controlling for the covariates (socio-economic status, cognitive skills, non-cognitive skills, educational attainment) reduces the correlation between communication and earnings by 80%. This chapter investigates the interaction between early communication and these broader predictors of earnings in adulthood. In particular, the question is “which covariates most explain differences in later earnings between children with low and high vocabulary, and in what ways are these different influences exerted?”.

Oaxaca (1973) and Blinder (1973) proposed an empirical method in the field of labour economics which can be applied to answer the following thought experiment: what proportion of an observed income gap can be explained by differences in the covariates, and what proportion is explained by differences in how those covariates are rewarded?

For example, what proportion of the income gap between low and high vocabulary individuals is due to the observed differences in educational attainment? Are the returns to higher education levels the same for low and high vocabulary groups?

In this context, for example, individuals with higher vocabulary scores at age five are likely to achieve higher educational attainment by the end of schooling. Higher levels of education are subsequently rewarded in the labour market, and this was controlled for in the previous analyses. The Oaxaca-Blinder Decomposition technique would show whether high levels of education are equally rewarded for individuals with different initial vocabulary scores, i.e., would targeting increased educational attainment in low vocabulary groups be likely to lead to increased earnings in later adulthood, or are the differences

The previous two analyses demonstrate a clear direct effect of communication ability on earnings. Typical development of speech and language between the age of two and seven predict higher household incomes at age 50, the only reliable measure available within Chapter 2. Higher receptive vocabulary scores at age five predict higher individual income across the ages of 30 to 42 in Chapter 3. These findings are in line with the literature available on accumulation of human capital in childhood, which commonly contains an element of language development (Almond et al., 2018; Conti et al., 2019).

This is of societal and political importance. Earnings are not only important to the individual, predicting quality of life in various forms, but also to the government as they represent higher productivity, higher tax returns, reduced propensity for the individual to commit crime (Becker, 1968; Fu & Wolpin, 2013) and reduced usage of health care services (Mangalore, 2006; Hirsch, 2013). This is discussed further in Section 1.3.5.2 of the literature review.

The direct focus on communication skills within this thesis represents a novel contribution to the literature on early human capital predicting later life outcomes, particularly income. However, the lack of prior research means that there is currently little understanding of the interaction between communicative ability and the multiple other forms of relevant human capital which also predict income. For example, Lindqvist and Vestman (2011) found that non-cognitive traits were relatively more predictive of unemployment and earnings within the lowest earners, whilst cognitive traits were relatively stronger predictors of earnings within skilled occupations. The ability to understand and communicate with others represents elements of both cognitive and non-cognitive skills but had not previously been individually assessed for predicting earnings.

Within Chapters 1 and 2, multiple covariates were controlled for in order to estimate the direct relationship between our verbal communication variable and our outcomes. An unadjusted model in Chapter 3 would report a coefficient around five times as large (see Appendix Figure A3.6) in equivalent samples. Whilst controlling for covariates is necessary to provide a reliable direct effect estimate, the interactions between language and the associated covariates may be vital in understanding the complete relationship between language and later life earnings. Furthermore, these interactions may further our understanding of the compounding nature of various forms of human capital explaining the general correlations between cognitive and non-cognitive traits.

Communication as a skill is central to both internal cognitive processing and external relating to others. Within the context of cognitive and non-cognitive human capital, communication could be the skill which enables these other traits to cluster within an individual. Alternatively, communication could affect the value of other skills within the labour market. For example, strong mathematical ability is valued by many companies but the ability to communicate mathematical models effectively to colleagues or clients may enhance their overall effect on workplace productivity which determines wages (Lazear, 1981). Alternatively, being able to effectively express one's value to the company may similarly improve the returns to other valuable traits through salary negotiations. There are many hypothetical ways in which communication may enable an individual to increase the recognised value of a given set of skills.

Within the broader branch of communication skills, discussed in the literature review, the BCS70 data captures receptive language, a highly cognitive element of communication. The above considerations show that it is beneficial to our understanding of the importance of language to be flexible and diverse in the way we model its multiple effects. The above considerations require an alternative approach to measuring the effect of language using a form of moderation analysis.

This chapter employs a decomposition technique to unveil some of the interactions between language, covariates and earnings.

First, we briefly summarise the literature relating to four key covariates, as identified through the literature reviewed in Chapter 1 and the statistical covariate selection process in Chapter 2. Each of these covariates relate to both early language development and later life earnings:

1. Parental SES
2. Cognitive Ability
3. Non-Cognitive Traits
4. Educational Attainment

Heterogenous effects are considered through restricted sampling by gender, for which the literature is also summarised in the following section. These literature summaries inform the specific research questions tackled by the decomposition technique.

4.2 Summary of correlates of both language and earnings

4.2.1 Parental SES

Socio-Economic Status, or SES, is a combination of parental characteristics usually containing parental occupational status, education level and income. Multiple risk factors relating to the level of social disadvantage within the family predict early language development within the child. Parental SES is commonly discussed in prevalence analyses of language disorders due to this correlation between social disadvantage and early child development (Dockrell et al., 2014; Christensen et al., 2017; Camacho et al., 2019; Law et al., 2017a). This section summarises the key findings relating parental SES to both child language and earnings, although further detail can be found within Chapter 1, Sections 1.2.5.1 and 1.3.5.2.

In particular, low parental education level is cited in each of these studies, either maternal, paternal or both, as being a key risk factor for poor language development. Hart and Risley (1995) focus on paternal occupation, measured as white collar, blue collar or welfare, and the association with children's exposure to vocabulary. Similar studies have confirmed a gap in not just the number of words the child hears but the range of vocabulary the child is exposed to in parent-child interactions (Rowe, 2008; Hoff & Tian, 2005; Cartmill, 2016) or simply the parent's level of vocabulary (Parsons et al., 2011; Sullivan et al., 2017).

Parental vocabulary and parent-child interaction style are two common mechanisms used to explain the SES gap in early child language development. At the more extreme end of poverty, there is potential for a more causal relationship whereby the child may not have access to educational resources such as physical books or online resources (requiring internet access and a home computer regardless of whether the resource itself is free) due to the financial restraints on the family, although Shonkoff and Philips (2000) believe this to be only a small component of the observed differences. Time constraints upon the parents in lower skilled occupations may offer an additional explanation for the differences in parent-child interaction.

Whilst SLCN tends to occur more commonly in low SES groups (Basit et al., 2015), Reilly et al. (2014) find that being a 'late talker' is not any less common at age two within high SES households. It is important to clarify whether the associations between SES and language are present within our sample, and the measures identified in Chapter 3, in order to fully understand the associations with later life earnings.

SES has been previously described as "running in the family", such that parental occupational status and educational attainment often predicts the occupational status and educational attainment of the child (Björklund et al., 2007). Poorer children are considerably more likely to end up in poverty (Blanden and Gibbons, 2006). The persistence of income across generations, the absence of perfect social mobility, is an entire field of study in its own right (Solon, 1999; Blanden, 2015), for which we are currently interested in the role played by childhood vocabulary as a moderator.

Blanden and Machin (2010) explicitly tested the relative reduction in social mobility between 1958 and 1970 using the change in social gradients of a range of early cognitive and non-cognitive skills, including receptive vocabulary development. They found significant correlations between parental social gradients in their children's early human capital and later parental social gradients in the income of their children. In summary, the higher inequality observed in early skills, the greater inequality observed in subsequent economic performance. This shows that the link between parental SES and child vocabulary and the link between parental SES and their child's subsequent income as an adult are clearly connected and will impact on our findings when we measure the link between child vocabulary and subsequent child income.

4.2.2 Cognitive

Correlation matrices reveal that whilst individual specific cognitive abilities exist, there are overlaps in different functions and that an individual who excels in one area may be expected to perform well in others. Harvey and Miller (2017) find that receptive vocabulary in particular predicts mathematics scores in an incrementally valid manner to other Executive Function tasks. Table 1 in Lubinski (2004) demonstrate the correlation of cognitive ability across multiple domains including mathematics, reading and vocabulary. These three domains are directly measured within the BCS70 data used within this chapter, although cognitive associations measured within early years are discussed in detail within Section 1.2.5.3.

Within the BCS70, the correlations between various cognitive tests at age 5 are reported in Table 1 of Parsons et al. (2011), demonstrating that the Copying Designs Test, intended to measure visual-motor coordination (Rutter, et al. 1970), correlates most strongly to the other four tests and may therefore be the best measure of underlying transferrable cognitive ability. The English Picture Vocabulary Test (Brimer and Dunn, 1962) measure of receptive vocabulary also positively correlates to each of the other four cognitive measures reported.

Gough and Tunmer (1986) theorise that reading has two dimensions: decoding the alphabet and understanding the written word. The latter requires oral language abilities including vocabulary as the “lynchpin” (Zubrick et al., 2015).

Reading and vocabulary are intrinsically linked as cognitive abilities with reinforcing dual causality noted in the literature. This means that whilst oral language development precedes and predicts initial literacy (Law et al., 2017a) and vocabulary at age 4 predicts reading at age 6 (Muter et al., 2004), there is evidence that frequent and varied reading subsequently increases a child’s vocabulary (Newman and Dickinson, 2011; Pianta, 2012; Hattie, 2009). The closer the time-point at which language is measured to the time-point at which literacy is measured, the stronger the correlation. For example, low (defined as the bottom 15% of the distribution of scores on the Peabody Picture Vocabulary Test) language measured at age 5 would be expected to be a moderate risk predictor of low literacy at 10 (defined as the bottom 15% of the distribution of scores on the Academic Rating Scale literacy subscale) according to Zubrick et al. (2015). Low language at 8, by comparison, would have been categorised as a high risk of low literacy at 10. Associations between vocabulary at age 5 and literacy levels at age 34 have also been documented (Law et al., 2009) demonstrating the enduring nature of this relationship.

4.2.3 Non-Cognitive

Heckman and Kautz (2012) summarise a body of evidence demonstrating the value of non-cognitive skills, as defined in the broader literature review in Section 1.3.3, in determining earnings in later life. These “soft skills” are shown to explain an additional portion of the variance in earnings left unexplained by wage equations containing only cognitive measures of intelligence.

Borghans et al. (2008) find that children with low IQ scores often perform much closer to high IQ peers where incentives are higher, suggesting an underlying difference in non-cognitive skills. They hypothesise that different underlying traits that predict behaviour may therefore explain a portion of observed cognitive differences that may not be present in reality. A trait of applying maximum effort without requiring immediate incentive, characterised within

Conscientiousness and Emotional Stability, is desirable within the labour market and this may help to explain why cognitive scores in conjunction with non-cognitive scores collectively explain variance in earnings between individuals. These two forms of skills are therefore analysed separately within this chapter.

The "Heckman equation" uses a causal inference mechanism including time lagged cognitive and non-cognitive measures, i.e. they account for observations of these skills at multiple time points to control for the socio-economic endogeneity in early development of human capital. One of the key early childhood interventions analysed in the non-cognitive skills literature is the Perry Pre-School Programme, targeted at 3 to 4-year olds to increase IQ. This intervention explicitly references increased parent-child interactions as well as interactions between peers and teachers in the pre-school environment. The intervention improved IQ initially although these effects were not significant by age 8. Later life earnings, however, were significantly larger in the long run for participants explained by an increase in participant non-cognitive skills (Heckman et al., 2010). Improved externalising behaviour was one key mechanism used to explain the long run differences between those who participated in the intervention and those who didn't. The associations between both language and non-cognitive skills in childhood are implied, whilst the relationship between non-cognitive skills and earnings is empirically demonstrated. This chapter investigates the interactions between these factors.

Table 1.2 within Chapter 1 summarises all the current research associating non-cognitive skills to language in childhood. Chapter 2 demonstrates an association in later life between SLCN and two non-cognitive traits, openness to experience and extraversion, in a multi-variate model. Receptive vocabulary skills are mediated by literacy in their impact on externalising behaviour (Westrupp et al., 2020), whilst self-control has been shown to predict cognitive performance for children with language difficulties (McClelland et al., 2007). Chapter 3 did not find a significant mediation effect of locus of control in the relationship between receptive vocabulary and earnings.

The above research demonstrates the complexity in measuring the interactions between multiple forms of human capital. Language ability may be one factor which increases understanding of the overlaps between cognitive and non-cognitive skills. This chapter specifically focuses on the reasons behind the earnings gap in Chapter 3, and what role is played by these interacting covariates.

4.2.4 Educational Attainment

The above factors all suggest that language development is in some way a function of other early childhood factors, from socioeconomic advantage to more general cognitive and non-cognitive development. When predicting later life success using language development, the most important association in the literature is found in the education outcomes associated with early language development. Tomblin et al. (2008) argue that clinical services should be prioritised for those at risk of lower 'socially valued outcomes', primarily educational.

Parsons et al. (2011) demonstrate that poor vocabulary coupled with poor reading skills at age 5 and 10 within the BCS70 suggest a higher likelihood of leaving school at age 16 and not obtaining any years of non-compulsory education. Conti-Ramsden et al. (2010) similarly find that language and literacy skills predict qualifications achieved at the end of compulsory education after controlling for other elements of cognitive ability and family SES. Marsh et al. (2010) also report increased likelihood of obtaining 5 or more CSEs for children who receive intervention targeted at improving early language ability. The sample used for Marsh et al.'s (2010) analysis came from Boyle et al. (2007), where 162 children diagnosed with primary language impairment aged between 6 and 11 who received intervention with positive effects on their expressive and receptive language scores.

In traditional Mincerian models of income in later life, education is consistently a key feature (Mincer, 1962). Usually, this would be the total years of education or highest qualification achieved in models of the returns to education (see Psacharopoulos and Patrinos, 2018 for a full review) although data on the longer term educational attainment, i.e. beyond age 16, of children across the spectrum of early language abilities is lacking. Only Conti-Ramsden et al. (2018) and Chapter 2 provide estimates of the lower expected highest qualification obtained for an average individual with SLCN, relative to a typically developing peer.

Estimating the benefits in terms of education outcomes for children with poorer early language abilities requires controlling for the associated SES, cognitive and non-cognitive abilities described above as these are all similarly correlated with the same education outcomes (Chowdry et al., 2008; Jacob, 2002; Prevoo and ter Weel, 2015). However, failing to account for the expected differences in educational attainment between high and low performers on early vocabulary tests would miss a crucial part of the reasoning behind later observed income differences.

4.2.5 Gender

One final covariate of relevance to the relationships between language and earnings is gender. Boys have been found to be approximately twice as likely to be identified with language impairment as girls (Leonard, 2014), although Tomblin et al. (1997) found approximately equal proportions within longitudinal samples. Within the BCS70, males typically score around two points (approximately 0.2 standard deviations) higher than females on the English Picture Vocabulary Test, a finding discussed in Law et al. (2013).

Blanden (2006) found that receptive vocabulary at 5 and reading at age 10 in females were the strongest predictors of upwards social mobility, whereas in males the visual-motor co-ordination test of copying designs at age 5 was the strongest predictor. This shows again that male and female development and life experiences are likely to be significantly different.

Despite the relatively greater female engagement in higher education and similar overall cognitive and non-cognitive skills (although specific traits show gender differences), there still exists a gender pay gap such that males earn more than females on average (Costa Dias et al., 2018). Therefore, earnings equations are commonly estimated separately by gender to separately derive the returns to valued traits and target policies more effectively. One early application of this form of moderation analysis was the Oaxaca (1973) decomposition of the gender pay gap.

Groves (2005) also utilise the Oaxaca strategy of estimating separate samples by gender. Groves compares the estimated returns to a degree, the additional expected earnings as a result of obtaining a degree. They find that female earnings are more impacted by education level and that this helps to explain the current trend of increasing participation in higher education relatively amongst females. This is an example of gender being treated as a moderator of important factors in predicting earnings, such as education. Furthermore, Groves (2005) show that part of the difference in returns for a degree may be explained by different average levels of non-cognitive traits amongst females but also different average returns to non-cognitive traits amongst females. Non-cognitive traits have been shown to be valuable in determining academic outcomes (Heckman & Kautz, 2012).

For our analysis, we are already separating our sample into low, middle and high vocabulary groups. Therefore, for the initial analysis gender is included in the earnings equation as a dummy variable to account for the average gender pay gap. However, the secondary analysis further splits the sample by gender to estimate whether the relationship between vocabulary, the related covariates of SES, cognitive ability, non-cognitive ability and educational attainment, and subsequent earnings in adulthood is further moderated by gender.

4.3 Research Questions

The above literature and prior analyses suggest that vocabulary will predict the average levels of many covariates which predict an individual's earnings potential although this has not been previously analysed. Furthermore, there is an untested potential for vocabulary to moderate the effects of each of these covariates on earnings given the value of transferrable communication skills in the labour market. This literature leads to the following research questions targeted in the current analysis:

1. Are there significant differences in the covariates of low and high vocabulary groups?
2. Do these differences lead to an earnings gap between low and high vocabulary groups?
3. Does the vocabulary ability of the individual moderate the labour market value of these covariates?

4.4 Data

The data sample and variables are selected directly from the analysis conducted within Chapter 3. The complete case sample (n = 2,588) forms the basis for our earnings gap identified in Chapter 3. No new variables are introduced from the previous chapter's analysis. The covariates will be separated into the sections organised within the above literature review summary.

Vocabulary will be treated as low, medium and high (bottom, middle and top thirds of the distribution, separated by gender) to enable the Oaxaca-Blinder decomposition as our form of moderation analysis.

4.4.1 Receptive Vocabulary

The English Picture Vocabulary Test of receptive vocabulary at age 5 will be trichotomized into low, medium and high scoring groups. The bottom, middle and top thirds of the distribution for each gender are grouped. This is displayed in Figure 4.1.

Alternative specifications (bottom and top quartiles compared to the middle 50%, -1s.d. and +1s.d. compared to the middle 68% and rolling cut points) were considered. Thirds were used due to the benefit of equivalent sample size across the regressions compared in the decomposition, explained in detail within the methodology (Section 4.5). Each of the alternative specifications was also tested and results are compared in the Appendices.

Primarily, we will be analysing differences in adult income between the low vocabulary group and the middle or high vocabulary groups (two comparisons, low-middle and low-high) decomposed into average differences in each covariate and differential returns to each covariate.

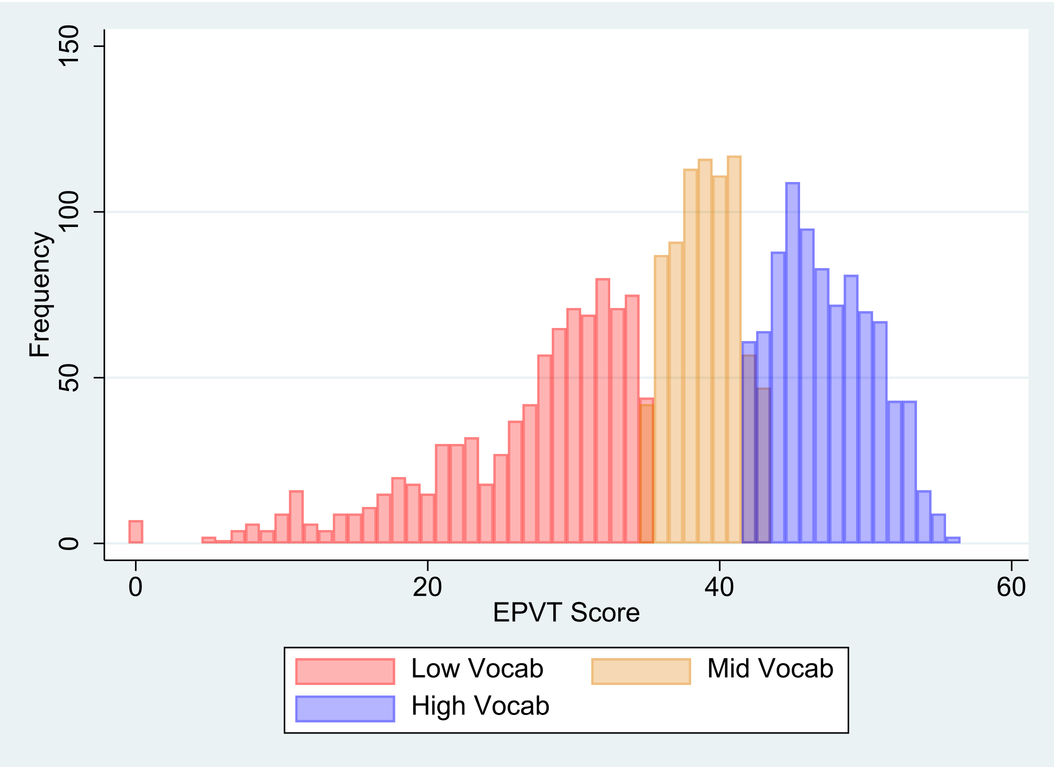


Figure 4-1 Distribution of EPVT Score by vocabulary group within complete case sample

Note: Slight overlap in cut-offs is due to different cut-offs for males (<=35 & <=43) and females (<=34 & <=41), due to slightly higher average scores amongst males. "Vocab" in the legend key is short for vocabulary.

The following sections summarise the outcome and covariates analysed, separated into the vocabulary groups depicted in Figure 4.1.

4.4.2 Outcome variable

The logged income of the individual, time-averaged across sweeps at age 30, 34, 38 and 42 will be used to represent that individual's earnings. The mean levels of income within each group is compared and referred to as the earnings gap.

Figure 4.2 shows that whilst high and low earnings are observed in every group, there is a mean level earnings gap between the groups. There is also a general trend that the most common low earners are within the low vocabulary group and vice versa.

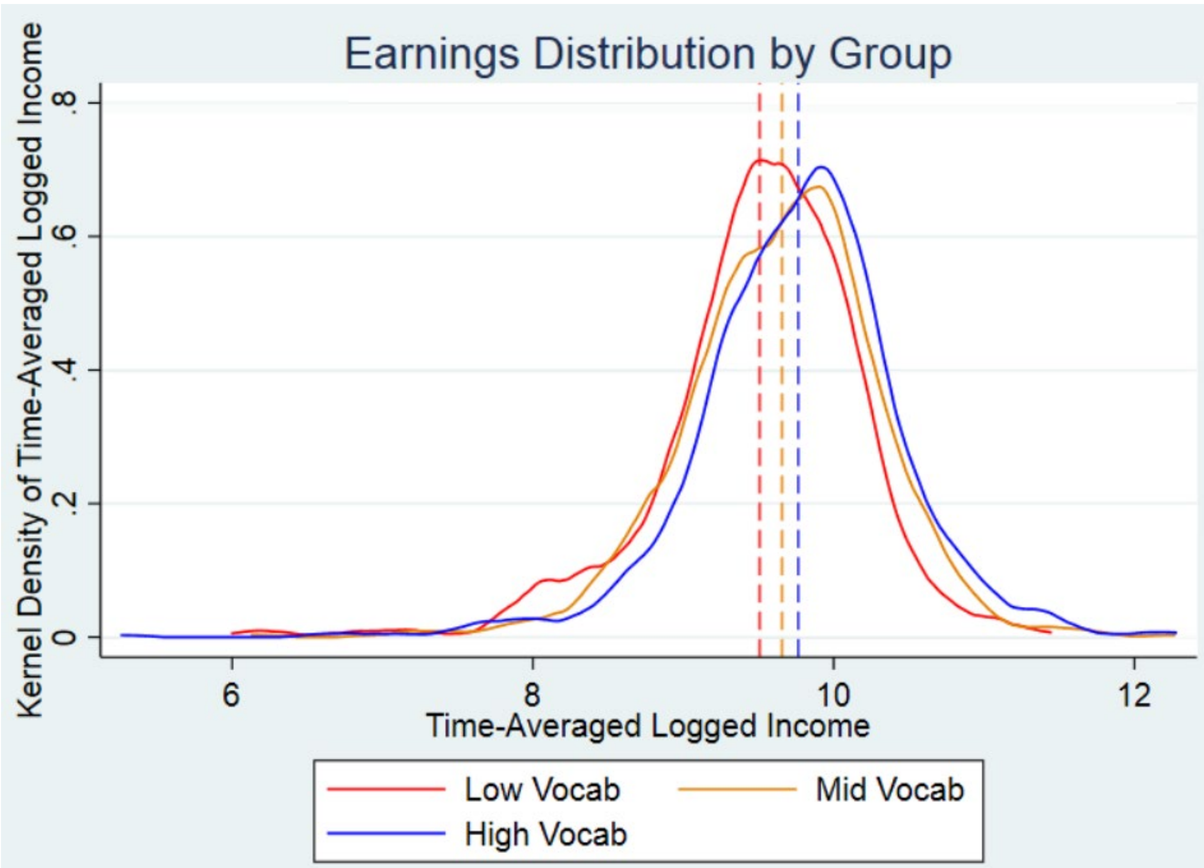


Figure 4-2 Distribution of logged time-average earnings by vocabulary group

Note: Kernel distribution plotted, with mean values to be compared for each group plotted as dashed vertical lines.

4.4.3 SES Covariates

Figures 4.3 to 4.6 depict the education levels of both the cohort member's parents separated by vocabulary group. Education level is split into CASMIN levels (level 2 at ≥ 5 CSEs, level 3 if a degree is obtained) as in the previous analysis. Figure 4.7 represents the registrar general occupational class of the father at birth separated by vocabulary group. Clearly, each of these SES characteristics are correlated with the child's vocabulary group (low, middle and high) at age five.

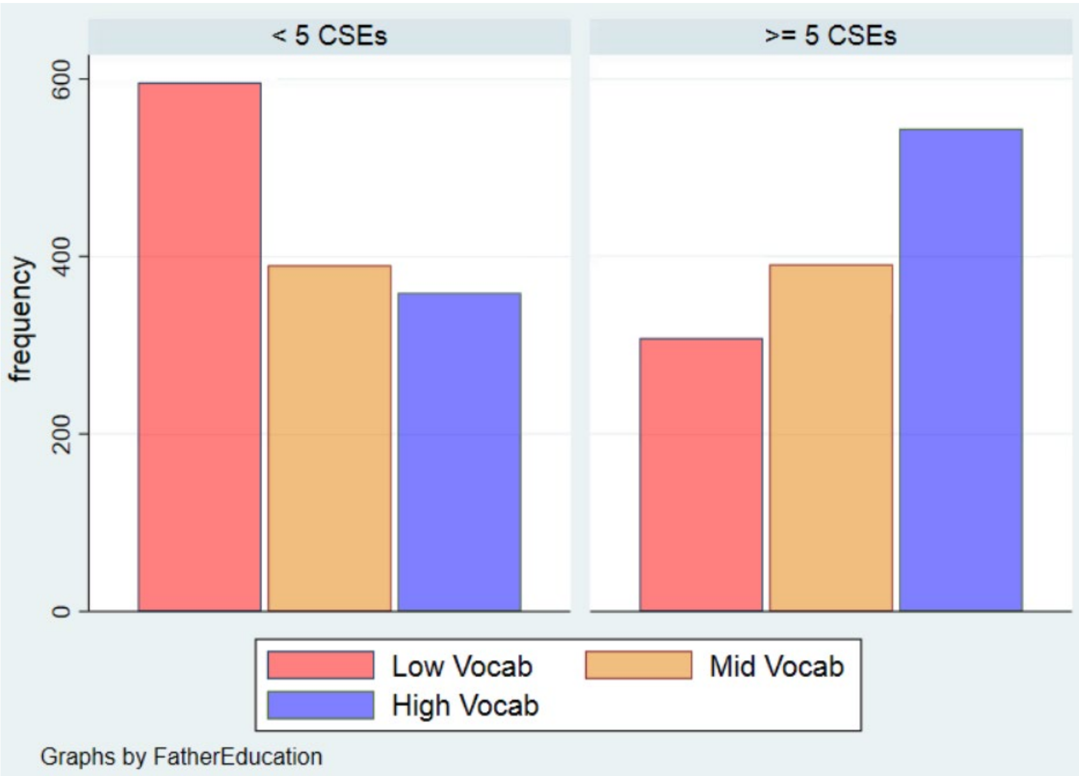


Figure 4-3 Father's education (CASMIN Level 2) distributions by vocabulary group

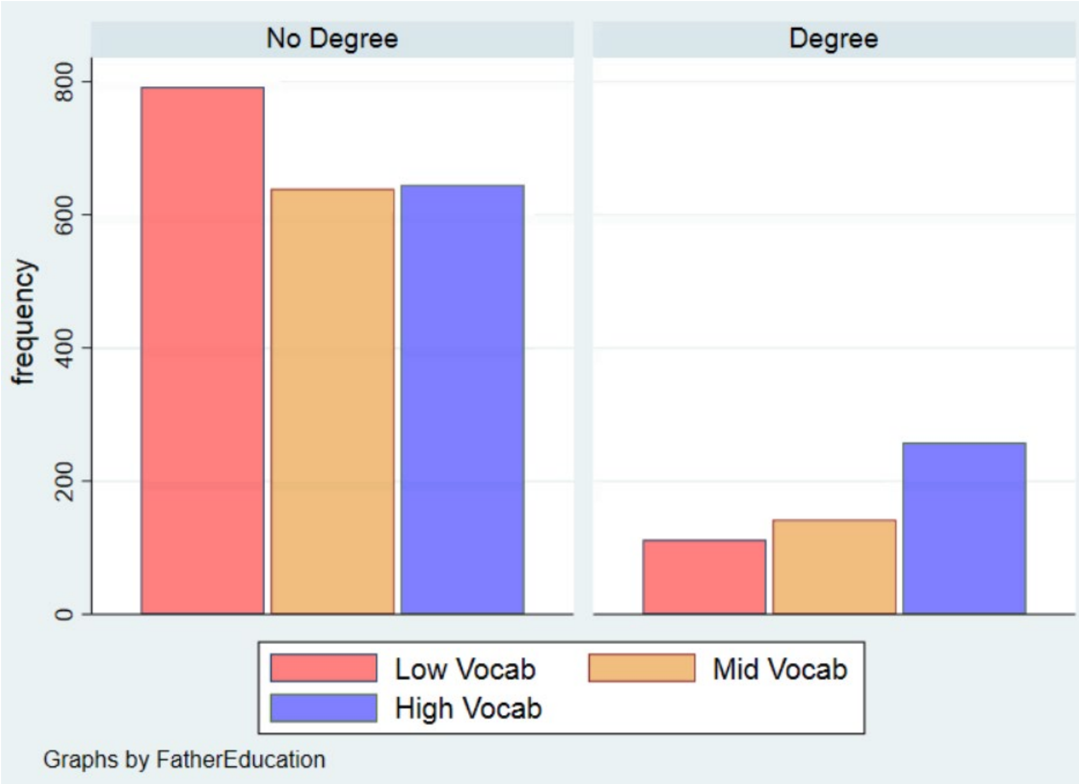


Figure 4-4 Father's education (CASMIN Level 3) distributions by vocabulary group

Note: Y axis plots frequency, not proportion, and relatively few parents achieved undergraduate degrees.

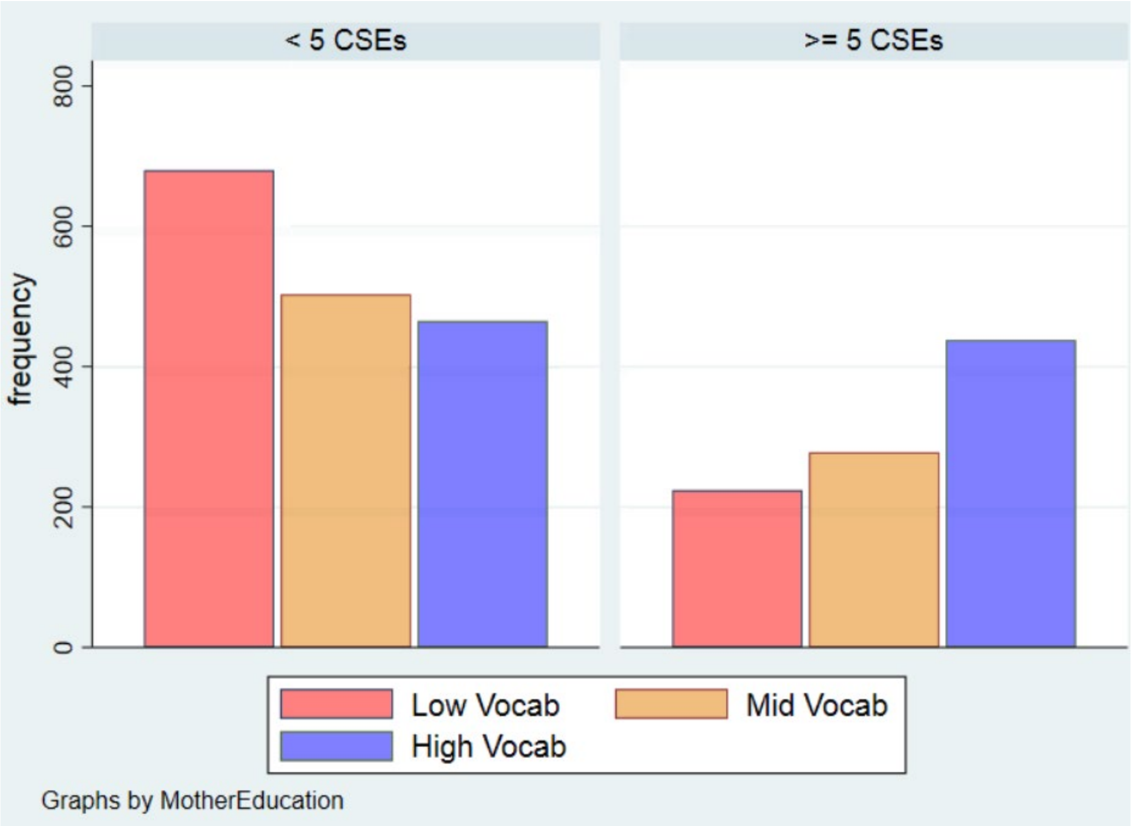


Figure 4-5 Mother's education (CASMIN Level 2) distributions by vocabulary group

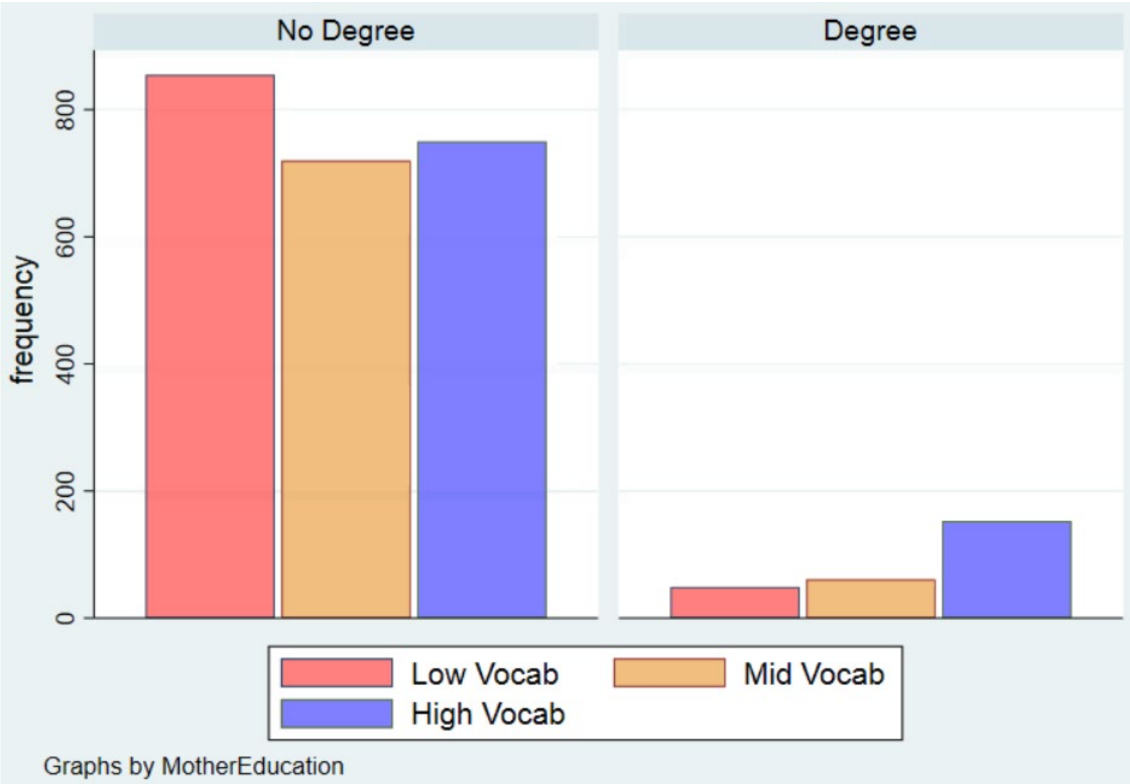


Figure 4-6 Mother's education (CASMIN Level 2) distributions by vocabulary group

Note: Y axis plots frequency, not proportion, and relatively few parents achieved undergraduate degrees.

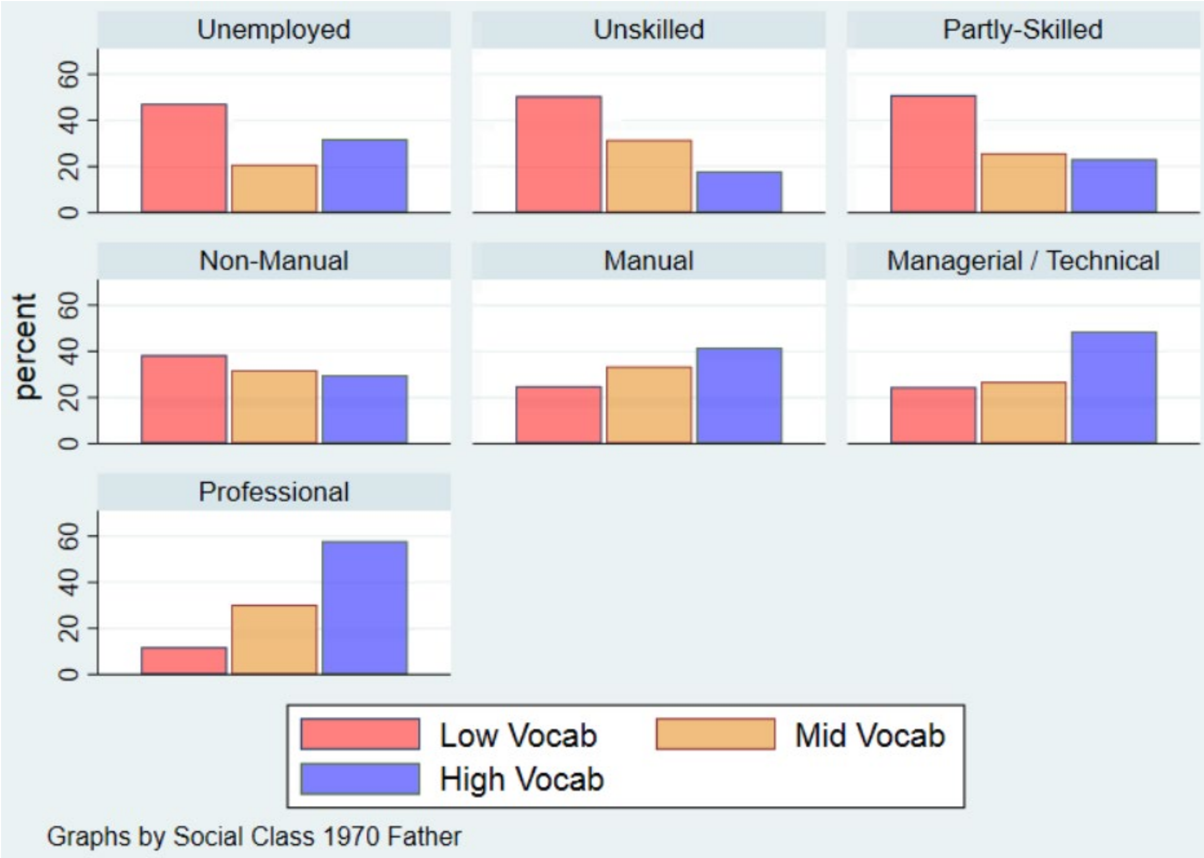


Figure 4-7 Vocabulary group distributions within each fathers' occupational social class category

Note: Mother's occupation is used where the father's is missing. Seven categories used, the six registrar social classes in 1970, and an unemployed category.

4.4.4 Cognitive Covariates

Figures 4.8 to 4.10 depict the distributions of copying designs score at 5, reading scores at 10 and maths scores at 10 by vocabulary group. As expected, higher vocabulary correlates with higher mean values of each of the other cognitive measures.

4.4.4.1 Figure 8

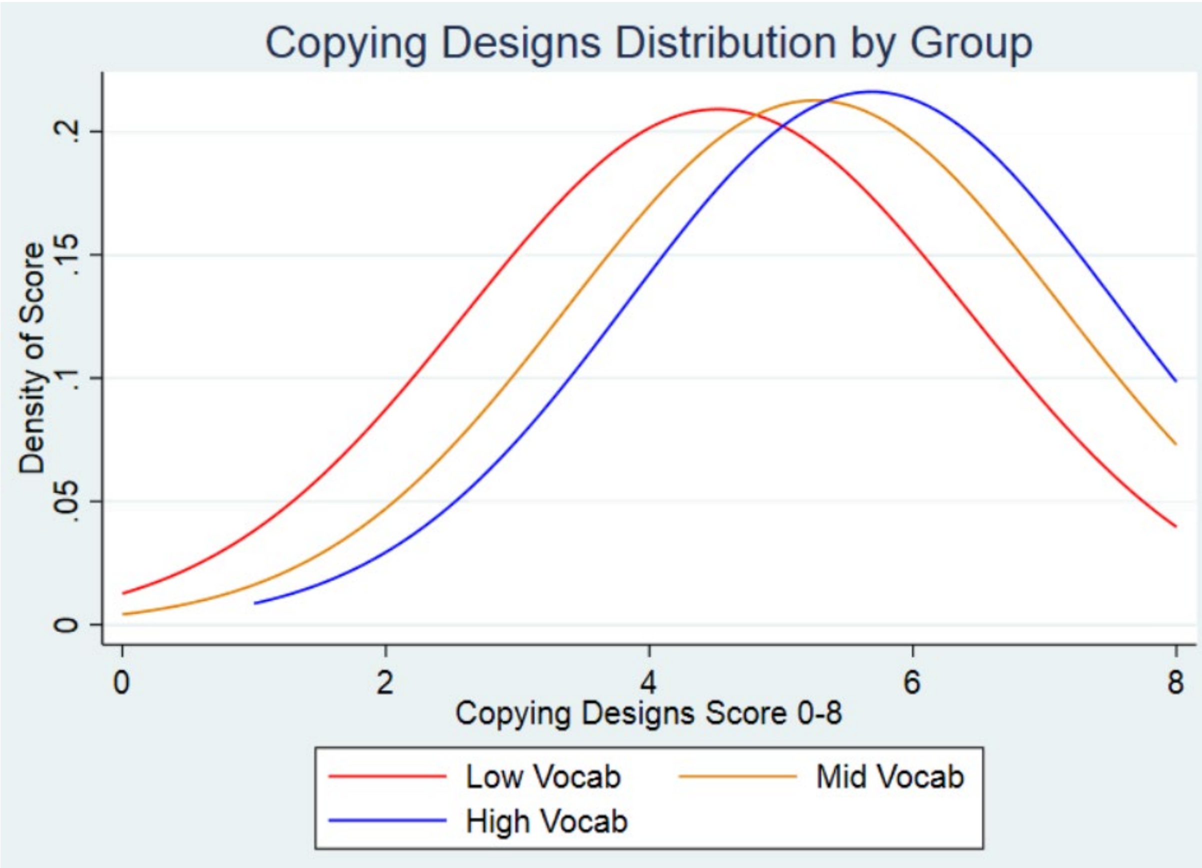


Figure 4-8 Distribution of copying designs test score by vocabulary group

Note: Normal distribution is fitted to scores due to the low range of discrete scores range from 0-8. Kernel density distribution would plot in a peak at every score, whereas the normal distribution shown highlight the different mean values within each group at the distributional peaks. Higher scores representing better visual-motor co-ordination at age 5, as is strongly correlated to other cognitive measures at age five (Parsons et al., 2011).

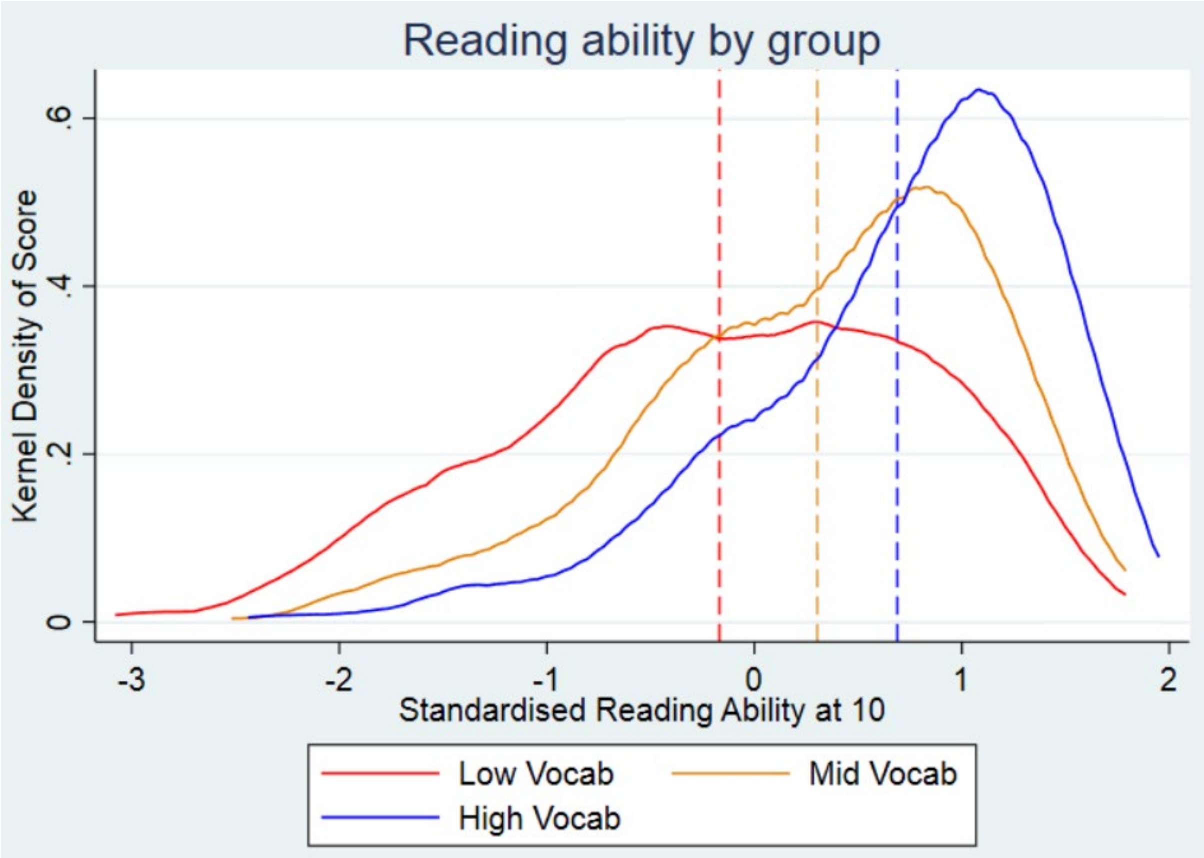


Figure 4-9 Distribution of reading scores by vocabulary group

Note: Kernel distribution plotted, with mean values to be compared for each group plotted as dashed vertical lines. Zero is the overall sample mean as this score is standardised.

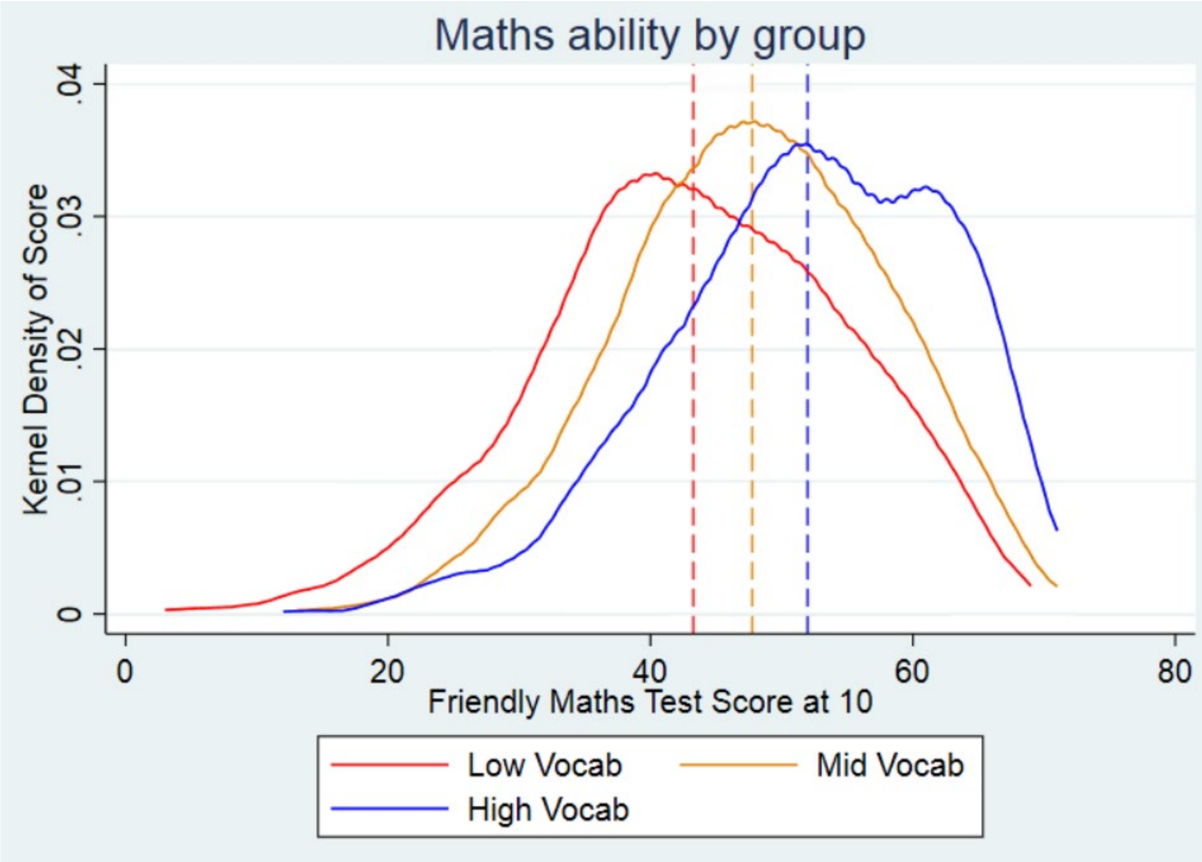


Figure 4-10 Distribution of maths scores by vocabulary group

Note: Kernel distribution plotted, with mean values by group displayed as dashed vertical lines.

4.4.5 Non-Cognitive Covariates

Conscientiousness and Locus of Control (defined in previous analysis) are plotted as our non-cognitive predictors of income. The association with vocabulary is weaker than that observed with the cognitive or parental SES measures. Whether this is a consequence of the measures themselves (discussed in previous analysis) or a genuinely lower weaker association is unclear. However, there is still a trend towards higher non-cognitive skills within the highest vocabulary groups in Figures 4.11 and 4.12.

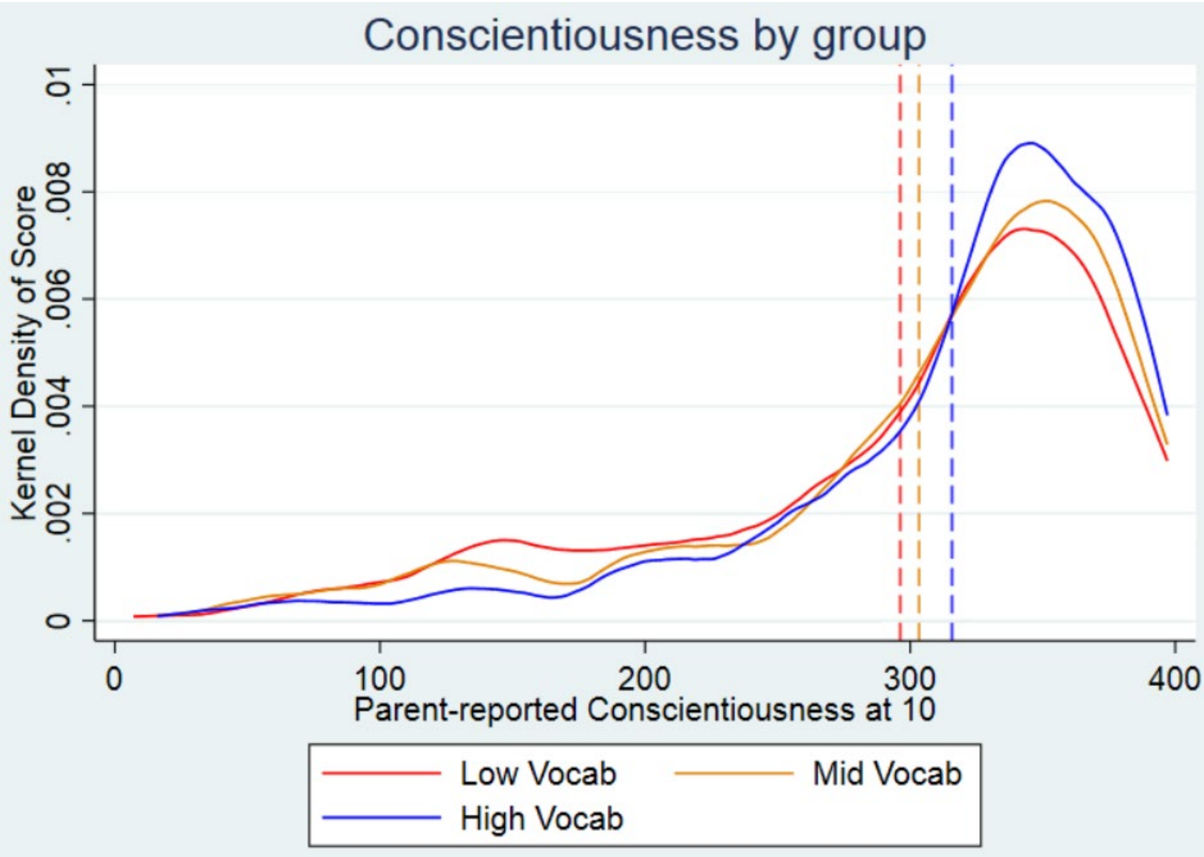


Figure 4-11 Distribution of conscientiousness by vocabulary group

Note: Kernel distribution plotted, with mean values by vocabulary group displayed as dashed vertical lines.

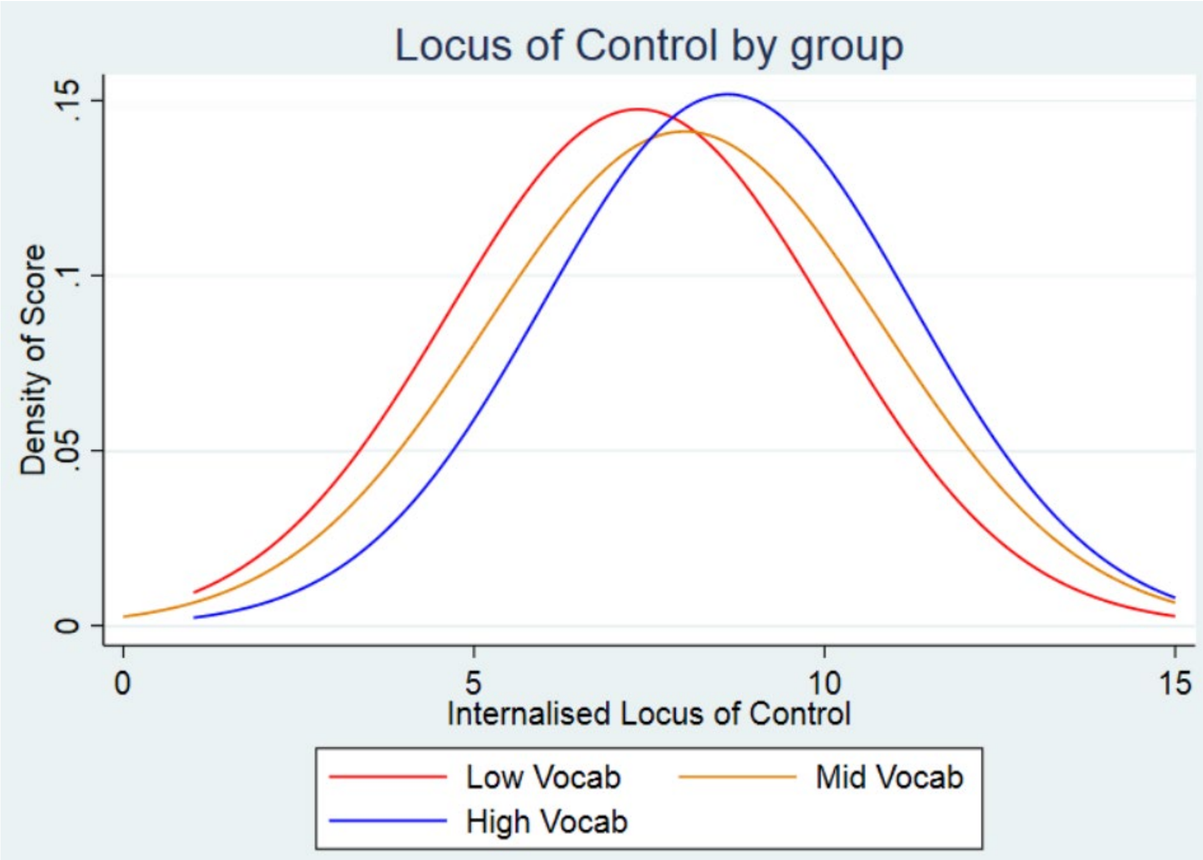


Figure 4-12 Distribution of locus of control by vocabulary group

Note: Normal distribution plotted due to the heaping of kernel density plots associated with a low range of discrete scores. Therefore, the peak of each curve represents the mean value within group. Higher locus of control represents more internalised locus of control, or a higher self-belief in control over one's own outcomes.

4.4.6 Educational Covariates

As expected, the higher vocabulary groups achieved greater educational qualifications at the end of compulsory schooling and were more likely to continue into higher education. These associations are displayed in Figures 4.13 to 4.15. It is notable how many of the low vocabulary group achieved no formal academic qualifications at the end of compulsory education or by age 30 in Figure 4.15. It is also clear how much more likely an individual with high vocabulary skills at age 5 is to obtain a degree level qualification which is known to be valued in the labour market (Psacharopoulos and Patrinos, 2018).

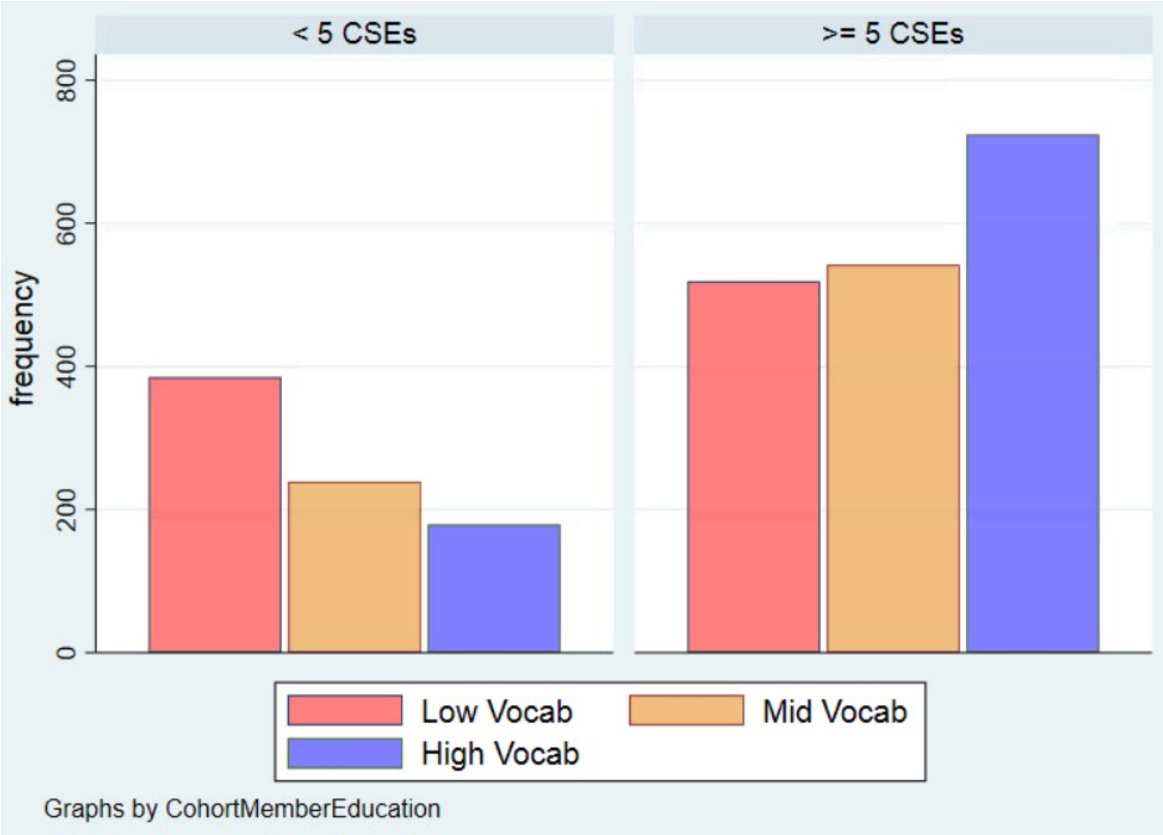


Figure 4-13 Distribution of CASMIN Level 2 at age 30 by vocabulary group

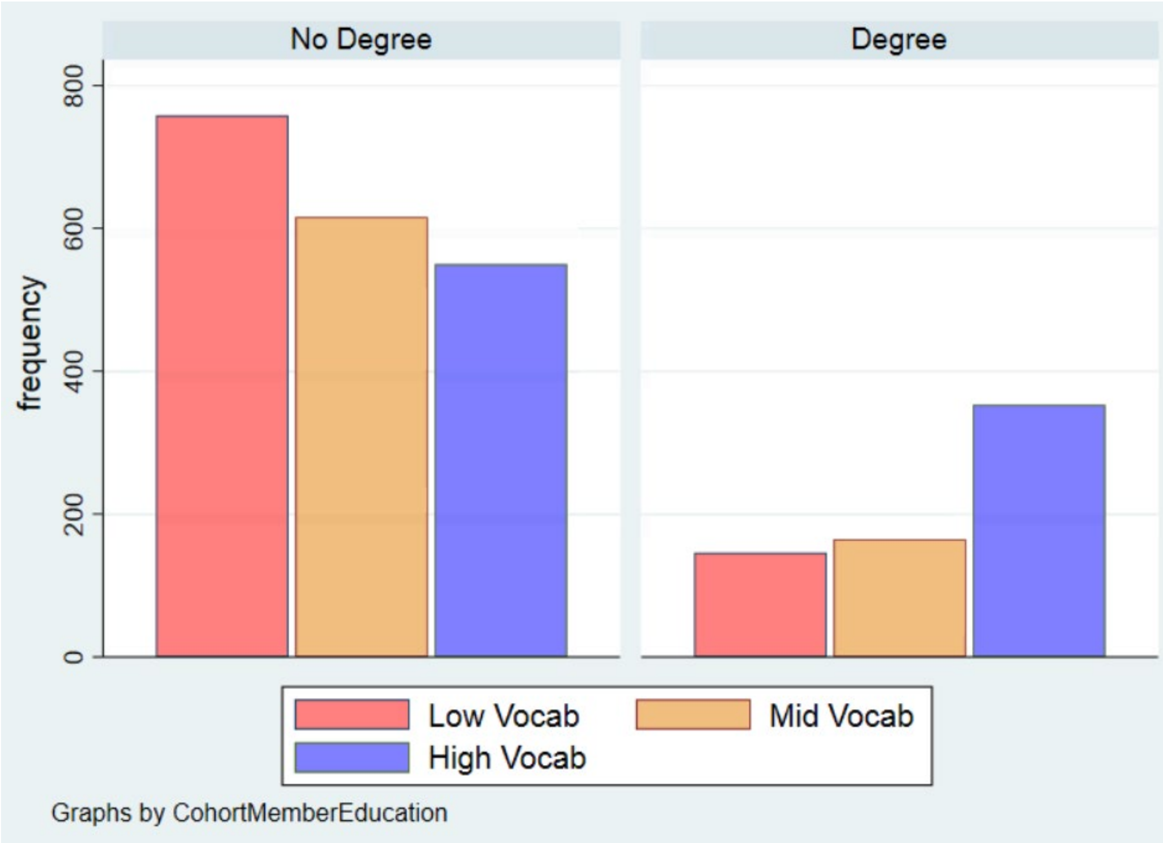


Figure 4-14 Distribution of CASMIN Level 3 at age 30 by vocabulary group

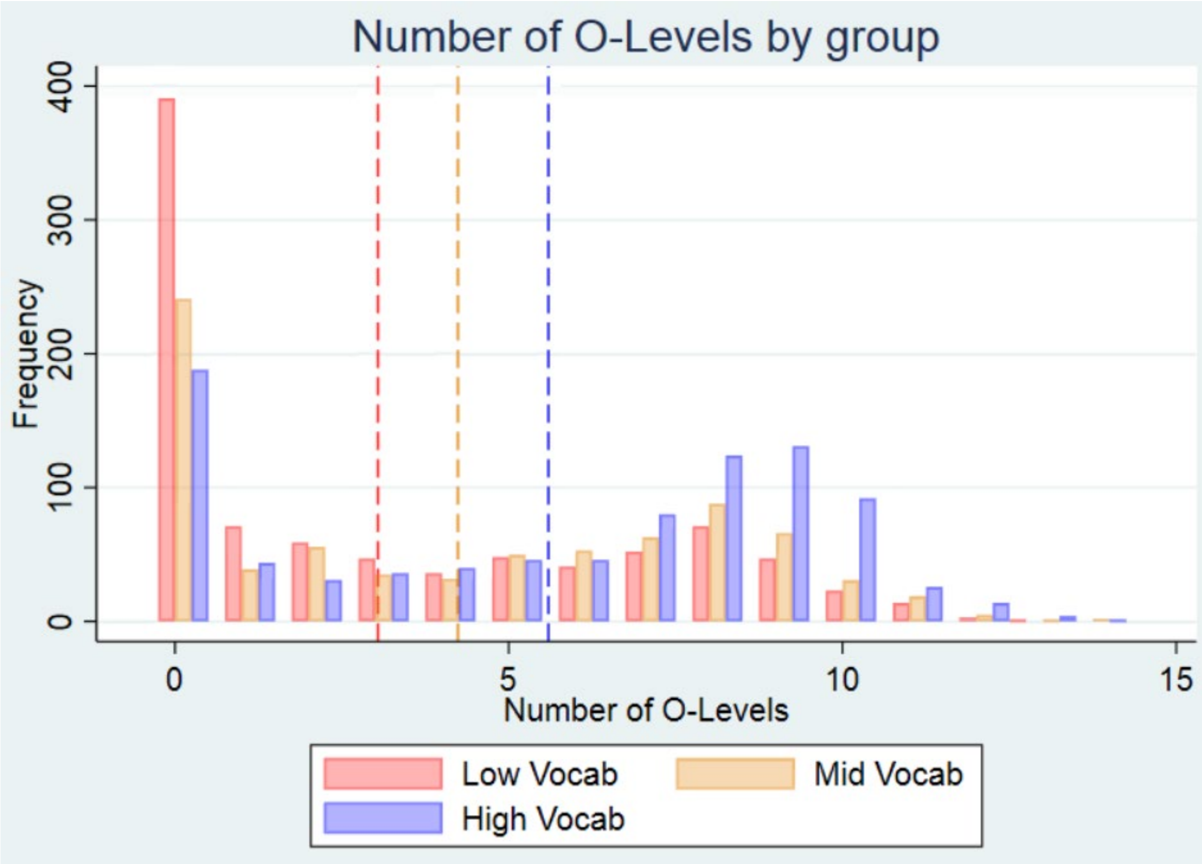


Figure 4-15 Distribution of the number of secondary education level qualifications by vocabulary group

Note: No account is made of the number of exams sat, so this only represents qualifications achieved and not an indication of success rate, some cohort members may not have sat examinations and very few would sit 14. Dashed vertical lines represent group means.

4.4.7 Summary Statistics

Table 4.1 shows significant mean differences at the 1% level in almost every single covariate between the low vocabulary group and either of the other groups. The two exceptions were for conscientiousness and the cohort member's mother having a degree between the low and middle vocabulary groups, both of which showed significant Standardised Mean Differences at the 5% level.

This clearly illustrates the fact that vocabulary development is not randomly distributed. Early receptive vocabulary is preceded by parental influence, co-developed alongside other cognitive and non-cognitive measures and subsequently predicts educational attainment and income across the life course.

Table 4.2 shows the potential for multi-collinearity between variables in our analysis. The between-variable correlations are strongest when the variables are drawn from the same covariate type, as per the literature structure. For example, r=0.711 between Maths and Reading, the strongest correlation in the data drawn both from the cognitive covariates. By comparison, r=0.265 between Maths and Father's Occupational Class, the strongest correlation between maths and any other variable not from the cognitive covariates.

Therefore, multicollinearity was not considered to be a problem for estimating effects between separate covariate types, but it could lead to uncertainty in the β estimates of specific covariates from the same type. Our interpretations are therefore based upon grouped covariate results. For example, we may discuss the effect of cognitive covariates rather than specifically the effect of reading or maths. In the cases where more detailed interpretation is required, for instance when looking at the effect of a degree within the overall educational covariate selection, the correlations reported in the Table 4.2 can be considered as a measure of the reliability of our interpretation although the confidence intervals do account for this uncertainty.

Table 4-1 Mean group values of covariates and standardised mean differences between groups

Note: Standardised Mean Differences = SMD *p<0.1 = *, p<0.05 = **, p<0.01= ****

Variable Type	Variable	Vocab Group			Significant Differences	
		Low (1) μ	Mid (2) μ	High (3) μ	SMD of (2)-(1)	SMD of (3)-(1)
Outcome	Time-Averaged Logged Income	9.507	9.656	9.765	0.215***	0.371***
SES	Father's Education CASMIN Level 2 (0/1)	0.217	0.319	0.317	0.237***	0.233***
	Father's Education CASMIN Level 3 (0/1)	0.124	0.182	0.286	0.160***	0.447***
	Mother's Education CASMIN Level 2 (0/1)	0.194	0.278	0.316	0.205***	0.297***
	Mother's Education CASMIN Level 3 (0/1)	0.054	0.078	0.169	0.090**	0.436***
	Father's Occupational Social Class	2.938	3.362	3.689	0.300***	0.531***

	(0-6)					
	Copying Designs (0-8)	4.520	5.256	5.685	0.372***	0.589***
Cognitive	Standardised Reading Score	-0.168	0.304	0.689	0.471***	0.867***
	Friendly Maths Score (0-72)	43.272	47.757	51.975	0.364***	0.706***
Non-Cognitive	Conscientiousness Score (0-400)	296.219	303.341	315.746	0.080**	0.220***
	Locus of Control (0-15)	7.344	8.000	8.617	0.234***	0.454***
Education	CASMIN Level 2 (0/1)	0.574	0.694	0.802	0.250***	0.476***
	CASMIN Level 3 (0/1)	0.162	0.211	0.391	0.126***	0.579***
	Number of O-Levels (0-13)	3.040	4.239	5.594	0.313***	0.667***

Table 4-2 Summary of complete case sample linear correlations between covariates by covariate type

*Note: Weakest and strongest Pearson's linear correlations within and between covariate types reported.   p<0.1 = *, p<0.05 = **, p<0.01= ****

		SES	Cognitive	Non-Cognitive	Education
SES	Weakest correlation	**0.049**** Father has >=5 CSEs: Mother has a degree	-	-	-
	Strongest Correlation	**0.414***** Father's Occupation: Father has a Degree			
Cognitive	Weakest Correlation	**0.0625***** Copying Designs: Father has >=5 CSEs	**0.385***** Copying: Reading	-	-
	Strongest Correlation	**0.265***** Maths: Father's Occupation	**0.711***** Maths: Reading		

		SES	Cognitive	Non-Cognitive	Education
Non-Cognitive	Weakest Correlation	**0.006** Conscientiousness: Father has >=5 CSEs	**0.161*** ** Conscientiousness: Copying Designs	**0.171*** ** Conscientiousness: Locus of Control *(only one correlation to measure)*	-
	Strongest Correlation	**0.132*** ** Locus of Control: Father has a Degree	**0.404*** ** Locus of Control: Maths		
Education	Weakest Correlation	**0.054*** ** Cohort Member has a degree: Father has >= 5 CSEs	**0.258*** ** Cohort Member has a degree: Copying Designs	**0.166*** ** Conscientiousness: Cohort Member has a degree	**0.394*** ** Cohort Member >=5 CSEs: Cohort Member has a degree
	Strongest Correlation	**0.332*** ** Number of O-Levels of Cohort Member: Father has a degree	**0.530*** ** Number of O-Levels of Cohort Member: Reading	**0.305*** ** Number of O-Levels of Cohort Member: Locus of Control	**0.747*** ** Cohort Member >=5 CSEs: Number of O-Levels of Cohort Member

4.5 Methodology

4.5.1 Origins

Oaxaca (1973) and Blinder (1973) are jointly credited with the Oaxaca-Blinder decomposition. The Oaxaca-Blinder technique is a thought experiment based upon counter-factuals (states of the world which are not directly observed, but estimates are made). In this application, the decomposition estimates what a low vocabulary group would look like (in terms of earnings) if they were given the non-vocabulary attributes of the high vocabulary group.

This technique analyses differences in the earnings equations of two distinct groups to explain gaps in pay between groups. Originally, the two groups compared were split on gender (Oaxaca, 1973) and race (Blinder, 1973).

The technique uses the two regressions, one in each group, and then compares the expected outcomes of each regression (the predicted earnings from the data of each group). Differences in predicted earnings (the earnings gap) is then separated into three portions.

Firstly, a portion explained by different average levels of variables rewarded in the labour market, an “endowment effect”. In this instance, our endowment effect will relate to the higher average levels of covariates in the higher vocabulary groups, as depicted in the prior section. For example, how much of the earnings gap between low and high vocabulary groups would remain if the low vocabulary groups were attributed equal levels of socio-economic status to the high vocabulary group.

Secondly, a portion of the earnings gap is explained by different returns to these covariates between the two groups (gender or race in the initial models), a “coefficient effect”. For example, how much of the earnings gap between low and middle vocabulary groups would remain if the low vocabulary group were rewarded at the same rate for their maths ability as the middle vocabulary group.

Finally, there is a portion known as the “interaction effect” which accounts for the fact that the endowments and coefficients may vary between groups simultaneously. This term is necessary to allow accurate estimation of the prior two effects, although rarely interpreted as meaningful. Masters (1974) characterises the interaction effect as any additional disadvantage experienced by an individual in the disadvantaged group who has high levels of the covariates. Despite achieving the beneficial covariates, they are not rewarded for these at an equivalent rate if an interaction effect is present.

The decomposition relies upon a chosen perspective of one of the two groups, usually the disadvantaged group with the lower earnings. For us, this will be the low receptive vocabulary group identified as the bottom third of the distribution at age 5 (see Figure 4.1). Oaxaca (1973) defined the earnings gap that would exist in the absence of discrimination as the endowment portion of the earnings gap, as any other portion must be due to differential treatment between groups. However, this definition of discrimination is contested and discrimination is not the focus of this study; there are clearly vocabulary differences between the two groups that could explain why returns to specific skills may be higher within the high vocabulary group.

The following section outlines the statistical derivation of the decomposition, as this is useful in demonstrating its later application and interpretation.

4.5.2 Derivation

Firstly, take the equation of Income in Group A. Usually, A is considered to be the higher-earning, advantaged group but this is not essential.

$$\ln(Y_{Ai}) = \beta_{Ai,k} X_{Ai,k} + \varepsilon_{Ai}$$

The logged wage of an individual, 'i', in Group 'A' is predicted by 'k' number of variables 'X'.

For example, X could be "Experience", "Education Level" and/or "Hours worked" for each individual in the Group.

β would represent the additional earnings for every unit increase in a given X in this model. For example, β_1 would represent the expected average increase in income predicted by an extra unit (year) of experience for an individual in Group A. β_2 would represent the expected average increase in income predicted by an extra unit of Education (e.g. a year or a qualification) for an individual in Group A.

Within these β parameters there would be a constant term, previously denoted α in regression models, representing the average earnings of an individual with zero amount of all predictors.

Finally, there is the residual error term, 'ε_i', which shows the distance between the true logged income of the individual and the logged income predicted by the model. The β parameters estimated by Ordinary Least Squares (OLS) minimise the sum of the squared residual errors to provide a line of best fit for our data.

If we were to simply predict the average income in Group A, we would need to know our β estimates from our regression and our average level of every X variable.

The error term would be zero on average by definition as the OLS linear regression models used do not systematically under or overpredict within a sample (although this inference is not necessarily true of the estimates extrapolated for the broader population).

This gives the following equation for the average earnings in Group A:

$$\text{mean}(\ln(Y_A)) = \beta_{Ai,k} \quad * mean(X_{Ai,k})$$

Or, if we were to simplify k to just one predictive variable for the derivation:

$$\text{mean}(\ln(Y_A)) = \beta_A * mean(X_A)$$

In short, if only 1 variable predicted earnings, we only need to know the average level of that variable and the returns to it in order to predict the average level of income.

For example, say X = Hours worked, and Y = weekly pay. If we know that individuals in Group A work 40 hours per week on average, then we only need to know our OLS estimate of the hourly rate of pay to work out the weekly rate of pay.

This method is robust to as many X variables as the model requires (Oaxaca, 1973).

We conduct exactly the same analysis within Group B. As these are a mutually exclusive group to Group A, their levels of every X variable and their level of income represent a totally new set of sample observations. Therefore, we will obtain a totally different set of estimated parameters, or β's. We refer to the parameters as β_B's to represent the second group and regression.

To show the average level of income in Group B:

$$\text{mean}(\ln(Y_B)) = \beta_{Ai,k} \quad * mean(X_{Bi,k})$$

Or, if we were to simplify k to just 1 predictive variable for the derivation:

$$\text{mean}(\ln(Y_B)) = \beta_B * mean(X_B)$$

We are analysing these two groups separately as we believe there is a difference in income between the two groups. On average, that difference in income is shown as:

$$Average\ Income\ Gap = \text{mean}(\ln(Y_A)) - \text{mean}(\ln(Y_B))$$

Or in our model with one predictor variable:

$$Average\ Income\ Gap = \beta_A * mean(X_A) - \beta_B * mean(X_B)$$

What we want to do is to separate the Income Gap into that which is due to simply the differences in the mean levels of the predictor, and what is due to differences in the parameters. To do this, we keep the same equation, but we add in a number of “zero” terms, partially based upon “counter-factuals”, estimates of what would happen in alternative worlds. These terms will estimate the earnings of an individual in Group B if they were in Group A and vice versa. Each of our zero terms will be in a set of square brackets emboldened initially to show that the equation has not actually changed.

$$= \beta_A * mean(X_A) - \beta_B * mean(X_B)$$

$$+ [\boldsymbol{\beta_B * mean(X_B) - \beta_B * mean(X_B)}]$$

$$+ [\boldsymbol{\beta_A * mean(X_B) - \beta_A * mean(X_B)}]$$

$$+ [\boldsymbol{\beta_B * mean(X_A) - \beta_B * mean(X_A)}]$$

When removing the square brackets to have an expression with 8 terms, we can begin to factorise into the portions which will make up our decomposition. If we were to number the 8 terms above from 1 to 8, such that line one shows terms 1 and 2, line two shows terms 3 and 4, line three shows terms 5 and 6 and the final line shows terms 7 and 8.

Firstly, we take out β_B as a factor of terms 7 and 4 to create what we will call our "Endowment" estimate. This is what Group B would earn if they had the same average level of the variable X as Group A. Blinder (1973) phrased this term as "the value of the advantage in endowments possessed by the high wage group *as evaluated by the low wage group's wage equation*". [3]

$$\text{Endowment} = \beta_B * [mean(X_A) - mean(X_B)]$$

Then we take X_B as a factor of terms 5 and 2 to create what we will call our "Coefficient" estimate. This is what Group B would earn if they were rewarded, i.e. had the same β parameters as Group A. Blinder (1973) phrased this term as "the difference between how the high wage group's wage equation *would value* the characteristics of the low wage group and how the low wage equation *actually values* them".

$$\text{Coefficient} = \text{mean}(X_B) * [\beta_A - \beta_B]$$

This leaves us with terms 1, 3, 6 and 8. These are the two positive real terms and the two negative counter-factual terms. We factorise these to represent the change in income that would occur if the two above changes occurred simultaneously, i.e. if the rate of returns to variable X and the mean level of variable X were simultaneously corrected between the two groups. Masters (1974) characterises this term as the discrimination which occurs to highly skilled individuals within the disadvantaged group, although Blinder (1973) says that this term has "no ready economic interpretation".

[3] Blinder (1973) evaluates from the perspective of the high wage group. Definitions are rephrased accordingly.

$$\text{Interaction} = \beta_A * mean(X_A) + \beta_B * mean(X_B) - \beta_A * mean(X_B) - \beta_B * mean(X_A)$$

$$= [\beta_A - \beta_B] * [mean(X_A) - mean(X_B)]$$

As the above equations contain all of our 8 terms summarising the average difference in Income:

$$Average\ Income\ Gap = Endowment + Coefficient + Interaction$$

$$= \beta_B * [mean(X_A) - mean(X_B)] + \text{mean}(X_B) * [\beta_A - \beta_B] + [\beta_A - \beta_B] * mean(X_A) - mean(X_B)]$$

$$Average\ Income\ Gap = Endowment + Coefficient + Interaction$$

$$= \beta_B * [mean(X_A) - mean(X_B)] + \text{mean}(X_B) * [\beta_A - \beta_B] + [\beta_A - \beta_B] * mean(X_A) - mean(X_B)]$$

The following section illustrates these terms using a simple example.

4.5.3 *Illustrative Hypothetical Example*

We observe two distinct groups, Group A and Group B. We observe only their hours worked and their incomes in a simple model. This model is self-created and only intended to demonstrate the power and inference of the decomposition.

In Group A, we observe that this sample work 40 hours a week on average and earn £600 per week. Our OLS regression estimates that this represents a £15 an hour wage rate.

In Group B, we observe that this sample work 20 hours per week on average and earn £100 per week. Our OLS regression estimates that this represents a £5 an hour wage rate.

$$\text{mean}(\ln(Y_A)) = \beta_A * mean(X_A)$$

$$600 = 15 * 40$$

$$\text{mean}(\ln(Y_B)) = \beta_B * mean(X_B)$$

$$100 = 5 * 20$$

Using our decomposition technique, we observe the following:

$$Average\ Income\ Gap = Endowment + Coefficient + Interaction$$

$$= \beta_B * [mean(X_A) - mean(X_B)] + \text{mean}(X_B) * [\beta_A - \beta_B] + [\beta_A - \beta_B] * mean(X_A) - mean(X_B)]$$

$$500 = 5 * [40 - 20] + 20 * [15 - 5] + [15 - 5] * [40 - 20]$$

Our endowment shows us that the Group B individuals would earn £5 per hour for an extra 20 hours per week if they worked at the average rate of Group A, explaining £100 of the total income gap.

$$Endowment = 5 * [40 - 20] = 100$$

Our coefficient shows us that the Group B individuals would earn £10 per hour more for each of their 20 hours worked if they were paid at the same rate as Group A, explaining £200 of the total income gap.

$$Coefficient = 20 * [15 - 5] = 200$$

Finally, our interaction shows us that if the Group B individuals were earning £10 extra per hour and worked 20 extra hours simultaneously then this would explain the remaining £200 of the income gap.

$$Interaction = [15 - 5] * [40 - 20] = 200$$

In summary, only £100 of the £500 income gap can be fairly explained by the mean level of X and the rest is in some way due to differential treatment between the two groups.

This model can be applied to each variable within a regression and they combine to give an overall Endowment, Coefficient and Interaction effect from the overall regression model.

Figure 4.16 shows this example graphically. The blue line represents the earnings equation for individuals in Group A, earning £15 per hour worked. There is no intercept as no earnings are predicted in the equation for those not working (this can be relaxed as explained later). The red line represents the equivalent equation for individuals in Group B who earn £5 per hour worked. There is a vertical line representing the mean hours worked for each group at 40 and 20 hours respectively. The total mean difference in earnings is represented on the Y axis as the £500 gap vertical gap between the intersections of each groups' mean hours worked within their own earnings equation.

If Group B were to be rewarded for the hours they do work at the rate of a Group A individual, they would shift at X=20 onto the blue line for Group A. This would result in a £200 vertical rise in earnings which is our Coefficient effect – The difference in earnings explained by differential treatment at Group B's endowment level.

If Group B were to increase their hours worked but not alter their wage rate, they would shift from their starting point from X=20 to X=40 along the red line, representing a £100 increase in earnings (20 hours at £5 per hour). This is our Endowment effect, labelled E. A smaller red line is shown to represent this shift as a cumulative effect on top of the Coefficient Effect.

Finally, the remaining portion is our Interaction effect representing an increase of 20 hours worked paid at the wage rate difference between the two groups (20 hours at £10 per hour extra) represented by I in our graph.

Figure 4-16 Hypothetical example data demonstrating the Oaxaca decomposition components

Note: Group A represent an advantaged group earning £15 per hour. They work 40 hours per week on average at this rate. Group B represent a disadvantaged group earning £5 per hour. They work 20 hours per week on average at this rate. The above graph demonstrates an Oaxaca-Blinder Decomposition of the overall earnings gap evaluated from the perspective of the disadvantaged group into the 3 components: Endowment Effect (E), Coefficient Effect (C) and Interaction Effect (I) using this fictional self-generated data.

As in Jones and Kelly (1984), there could also be a difference between the two groups' intercepts. This would represent the differential treatment of an individual with no endowments in each group. This was not factored into our hypothetical example data.

Within Figure 4.16, we can see that if the blue line were to shift upwards (an increase in the intercept with all else held constant), this would increase our Coefficient portion ("C" on Figure 4.16) of the earnings gap representing a differential treatment between groups regardless of endowments. The intercept term is discussed more in the Index Number Problem section but is simply a portion of the Coefficient effect attributed explicitly to the intercept difference.

4.5.4 Meaning of Treatment

This decomposition technique has traditionally been used to measure discrimination as it models differential treatment between groups. For the reasons detailed here, this application does not seek to measure discrimination.

Differential treatment between the two groups, shown primarily through the coefficient term but also the interaction, is often used to infer discrimination and has been used in court cases to this effect (Ashenfelter and Oaxaca, 1987). However, discrimination is not explicitly measured by the model if there are any unobserved variables that differ between the two groups. With our application, clearly vocabulary differs between the two groups. We are looking to compare the experiences of individuals from different vocabulary groups using their earnings equations. Discrimination is not empirically identified in this, or any, Oaxaca decomposition.

If Group A in our hypothetical illustrative example were more productive workers, then this would explain their higher wage rate and the coefficient effect would not be interpretable as discrimination. The decomposition design is quasi-experimental as opposed to causal, making it a contentious technique for analysis of the gender pay gap. The existence of any unobservable factor which predicts wages means that this decomposition will not effectively measure discrimination on the demand side of the labour market.

Furthermore, if discrimination exists which leads to different endowments then discrimination could similarly be underestimated by the decomposition. If Group B in our hypothetical example were forced to work no more than 20 hours per week, this would not be captured by the model. Their lower work hours are captured in the endowment effect, not the coefficient effect, and would therefore not be considered as discrimination. Within the gender pay gap literature, the returns to education (coefficient effect, used to infer discrimination) would not capture any systematically limited access to higher education reflected in the endowments of females. This is discrimination which impacts the supply side of the labour market and therefore appears in the endowment term.

For these reasons, we do not interpret our coefficient or interaction effects as "discrimination" effects. In our example, vocabulary is not included in our model but is used to separate the groups, so we are explicitly separating our groups on an intentionally unobserved factor. Instead, the coefficient and interaction effects will measure the significance of vocabulary as a moderator of other predictors.

4.5.5 Index Number Problem

The "Index Number Problem" examined in Oaxaca (1973) refers to the sensitivity of this decomposition to the reference group. In the simple hypothetical example examined above, we think from the perspective of the disadvantaged group and it is their mean endowment (X_B, as only a single variable was used to predict earnings) that is used to measure the coefficient effect of a change in returns. If we were to apply X_A to this equation, we would report a different value for our coefficient effect. Similarly, the endowment effect relies upon the β_B parameters but this could easily be reversed and measured using the β_A parameters.

If we were to reverse our perspective, the overall gap would be reversed although the Interaction effect would remain unchanged, leading to differences in our endowment and coefficient effects to compensate. Therefore, it is important to have a theoretical rationale for which perspective to analyse from. This thesis represents the first demonstration of an earnings gap explicitly predicted by vocabulary levels, limiting the available research from which to establish a reference group.

Generally, we think from the perspective of the disadvantaged group, as the intention is to make recommendations that would increase their earnings rather than reduce the returns for the advantaged group (i.e. the discrimination is negative to Group B, not positive to Group A). In the present application, it makes sense to continue with this tradition. We want to estimate how improving the vocabulary of a low vocabulary individual in isolation may moderate their returns to other factors (X_B and β_B) rather than how reducing the vocabulary of a high vocabulary individual would do the same.

Appendix Figure A4.2 reports the results of reversing the perspective although the primary results presented are those which make the most sense from the perspective of understanding the experiences of low vocabulary individuals.

Another solution to the reference group issue is to estimate using a pooled sample (including an indicator of group membership) in addition to the restricted samples. Coefficients and endowments are then compared to the group average, although this requires a separate derivation. Results are presented only as explained and unexplained components, from which less information is gained relating to the experiences of the different vocabulary groups. Appendix Figure A4.3 reports the two components (explained and unexplained) decomposition, as recommended in Jones and Kelly (1984), although conclusions are drawn from the primary analysis.

The other issue to consider methodologically is the sensitivity of the coefficient estimates, particularly the intercept, to the zero points within the data (Jones and Kelly, 1984). The intercept in any equation measures the expected outcome (earnings) if all the predictors (our five types of covariates) are equal to zero. What these zeroes mean is subject to interpretation.

For our analysis, we scale every continuous variable to begin at zero such that the intercept represents an individual with zero conscientiousness, mathematical ability, reading ability etc. relative to the rest of the sample. For our dummy variables, zero represents the absence of that beneficial factor such that the intercept would represent an individual without O-levels or a degree and parents without a degree or >= 5 CSEs.

Within our data, the Father's Occupation variable is difficult to scale. Figure 4.7 demonstrates the distribution of vocabulary between parents in each occupational category. The bottom category would be an unemployed father although the distribution of vocabulary shows that this is a worse indicator of a disadvantage than the unskilled employment category. This is potentially due to temporal unemployment of skilled parents at the timing of the survey. For this reason, we include seven variables in our analysis, one for each paternal occupational category and the individual effect of each group on earnings is estimated. The seven variables are grouped using the 'normalize' stata command within the 'oaxaca package' (Jann, 2008). This allows us to not select a reference category. Our intercept represents the "average occupation" for a father in each category. For this reason, that the zero value is not a zero value in the same sense as the other variables, we do not interpret our intercept coefficient effect as it is too sensitive to our SES coefficient effect containing parental occupational status.

Similarly, the dummy variable for gender requires one gender to be set as zero, and thereby this gender's treatment across groups is picked up by the intercept. The other gender would then have their differential treatment picked up by the coefficient effect on the dummy variable. As there is a larger earnings gap for females, when these are the reference category (dummy variable for males, such that the intercept reflects a female) the intercept coefficient effect explains a larger portion of the overall earnings gap which is compensated by the gender coefficient effect. As we are interested in whether the vocabulary moderates the experiences of males and females differently, we use a restricted sampling approach to discuss heterogeneous effects rather than rely upon the coefficient effect on gender reported within the pooled model which is sensitive to the reference (omitted) group.

The reason the zero choices matter is that they impact our intercept. Given that there is a fixed overall coefficient effect regardless of which categories are omitted, a change in the intercept must necessitate an equivalent change in the coefficient effect of the given categorical covariate. If, hypothetically, the intercept was larger when omitting the unskilled fathers rather than omitting the unemployed fathers (the earnings gap is larger within these groups) then we observe a reduction in the coefficient effect attributed to parental SES. This is interpreted as vocabulary moderating the impact of SES less, but is entirely an artefact of decisions about which group to omit.

In our simple hypothetical example, consider the decomposition if we were to set 20 hours per week as the 'zero' point and run the same decomposition. If $X_B = 0$ rather than 20, then our coefficient effect from hours worked becomes zero. Our coefficient effect at the intercept (now the y axis is at x=20 on Figure 4.16) becomes equivalent to the £200 no longer explained by the differential treatment of hours worked. In the context of multiple variables it is difficult to ascertain which poorly scaled zeros are being picked up by the intercept effect, but the overall coefficient effect remains a reliable interpretation of differential treatment between the groups.

For this reason, the majority of the robust interpretation will be carried out on the endowment effects, separated by covariate type, and on the overall coefficient effect. Interpreting the specific coefficient effects of each covariate type relies upon trust in the zero points reflected in the intercept and should therefore be treated with caution, although the estimates are reported in Table 4.3. This is consistent with the recent literature employing the Oaxaca-Blinder decomposition (Nguyen et al., 2019).

4.5.6 Standard Errors

The 'oaxaca' package in stata (Jann, 2008) reports standard errors for each effect compared between the two individual regressions. Bootstrapping was tested as an alternative method of producing standard errors with no significant differences observed (see Appendix Figure A4.4).

4.6 Model Specification

Three regressions (low, medium and high vocabulary) are estimated within each sample (pooled, females, males).

Initially, we estimate the earnings of the low vocabulary group using the 14 covariates established within the Data section (See Table 4.1) falling into the five covariate types (gender included as a variable when using the pooled sample) framed in the literature review summary sections. We then estimate the earnings of the middle vocabulary group using the same set of covariates with their higher mean levels relative to the low vocabulary group. We also estimate the earnings of the high vocabulary group who have the highest average covariate levels.

The Oaxaca-Blinder decomposition is applied to compare the low to the middle vocabulary group with results reported in Table 4.3 and plotted in Figures 4.17 and 4.18.

We do not compare the middle and high vocabulary groups directly. These groups are unlikely to be compared in any policy or research applications. The difference between the two models shown could be interpreted as representing this comparison. For example, the difference between the earnings gap from low to middle and from low to high would be the earnings gap from middle to high. Similarly, if the endowment effect of SES explains a similar difference in earnings in both models, this suggests that SES is similarly distributed between middle and high vocabulary groups. If the coefficient effect of Education explained a much larger earnings difference in the low to high vocabulary decomposition, this would suggest that the returns to education in the high vocabulary group were significantly different to those in not just the low vocabulary group, but also the middle vocabulary group.

Within these results, we can interpret the endowment effect of each variable confidently. We can interpret more confidently the variables with the least arbitrary zero scaling decisions. We cannot confidently interpret the intercept, gender coefficient effect or the SES coefficient effect in a meaningful way.

To estimate the potential moderating effects of vocabulary being impacted by gender, we then run the above analysis within the male-only sample and the female-only sample separately (Figures 4.19 and 4.20).

4.7 Results

As the results in Table 4.3 demonstrate, there is a significant pay gap between the lowest vocabulary group and the middle vocabulary group. There is a larger pay gap between the lowest vocabulary group and the high vocabulary group.

4.7.1 Research Questions

The results presented in our summary statistics as well as the following figures and tables answer our prior research questions as follows:

1. *Are there significant differences in the covariates of low and high vocabulary groups?*

Yes, shown in Table 4.1. Standardised mean differences between vocabulary groups are significant at the 1% level.

Standardised mean differences are all significant at the 1% or 5% level between low and middle vocabulary groups.

2. *Do these differences lead to an earnings gap between low and high vocabulary groups?*

Yes, shown by the significant endowment effect in Table 4.3. This is also demonstrated in Figure 4.17.

Around a third of the overall earnings gap between low and high vocabulary groups is attributed to the endowment effect of educational attainment. This effect is significant at the 1% level.

A further fifth of the overall earnings gap between low and high vocabulary groups is attributed to the endowment effect of cognitive skills, namely motor-visual skills, reading and maths.

Both the above findings extend to the comparison between low to middle vocabulary groups, although the overall earnings gap is reduced by the closer similarities of the groups in both language and endowment levels.

3. *Does the vocabulary ability of the individual moderate the labour market value of these covariates?*

Not significantly at the 95% level, although the coefficient and interaction effects in Table 4.3 are positive and account for 42% of the overall earnings gap between low and high vocabulary groups, despite being non-significant.

Similarly, there is no significant coefficient or interaction effect reported in Table 4.3 between the low and middle vocabulary groups.

4.7.1.1 Summary

The earnings gap between each group is primarily explained by differences in the mean levels of endowments across groups. In particular, endowment of other cognitive abilities and of success in the education system are significant with the educational endowments explaining over half the total endowment effect. This can be seen visually in Figure 4.18.

Table 4-3 Decomposition results for low vocabulary group

Note: Low vocabulary group is the reference group. Effects sum to explain total difference in logged earnings. Covariate-specific coefficients may be unreliable due to zero scaling decisions, but overall endowment, coefficient and interaction effects are robust. Continued overleaf.

*p<0.1 = *, p<0.05 = **, p<0.01 = ****

Decomposition	Middle Vocabulary	High Vocabulary
Comparison Group Earnings	9.656***	9.765***
Low Vocab Earnings	9.507***	9.507***
Difference in logged Earnings	0.149***	0.257***
Endowment Effect	0.087***	0.179***
Endowment Effect SES	0.017	0.0.19
Endowment Effect Cognitive	0.027**	0.049**
Endowment Effect Non-Cognitive	0.002	0.004
Endowment Effect Education	0.045***	0.117***
Endowment Effect Male	-0.003	-0.010
Coefficient Effect	0.042	0.053
Coefficient Effect SES	-0.043	-0.020
Coefficient Effect Cognitive	0.120	0.000
Coefficient Effect Non-Cognitive	-0.051	0.213
Coefficient Effect Education	-0.021	-0.050
Coefficient Effect Male	-0006	-0.034
Coefficient Effect Intercept	0.044	-0.056

Interaction Effect	0.020	0.025
Interaction Effect SES	0.004	0.012
Interaction Effect Cognitive	0.013	-0.001
Interaction Effect Non-Cognitive	0.002	0.023
Interaction Effect Education	0.001	-0.010
Interaction Effect Male	0.000	0.001

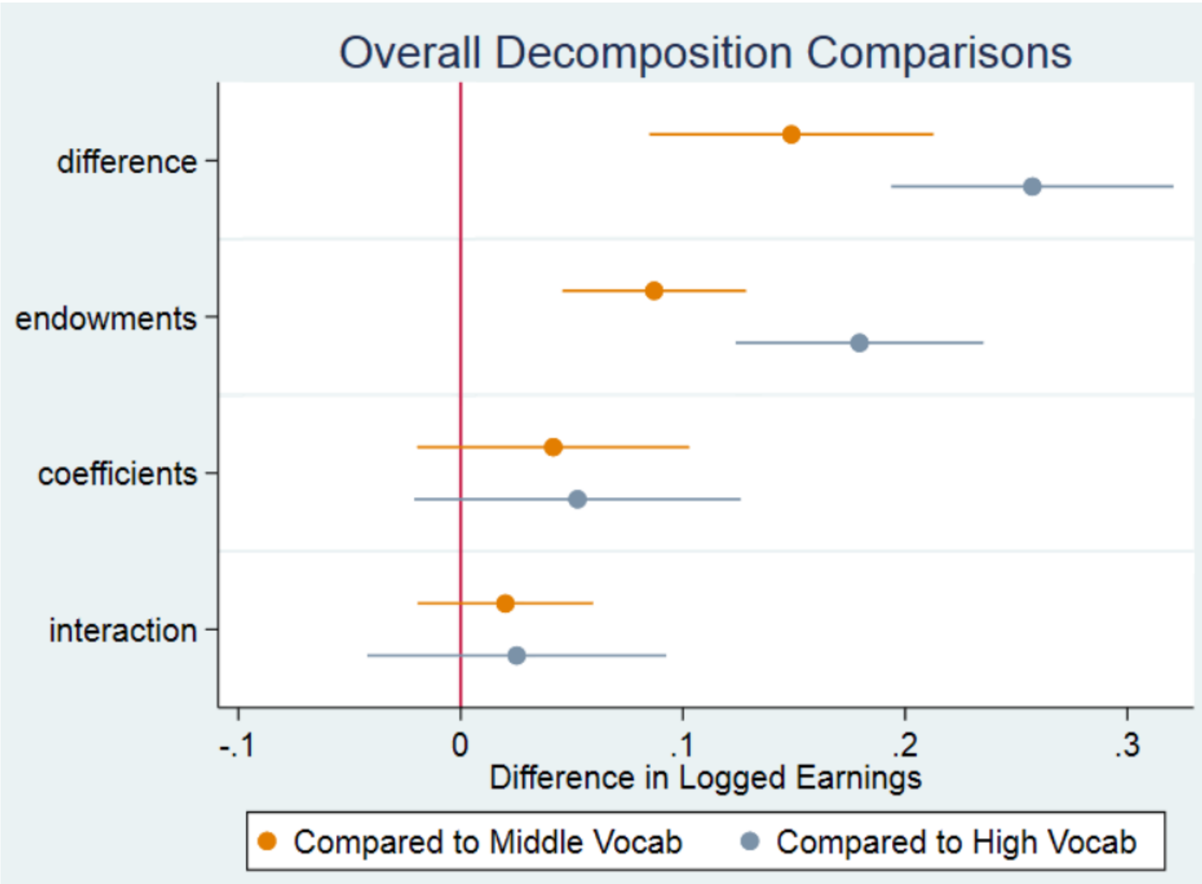


Figure 4-17 Endowment, coefficient and interaction components of overall earnings gap

Note: The low vocabulary group are our disadvantaged group in each derivation and are compared separately to the middle and to the high vocabulary group. Where the 95% confidence intervals cross the vertical zero line, point estimates are non-significant (5% level)

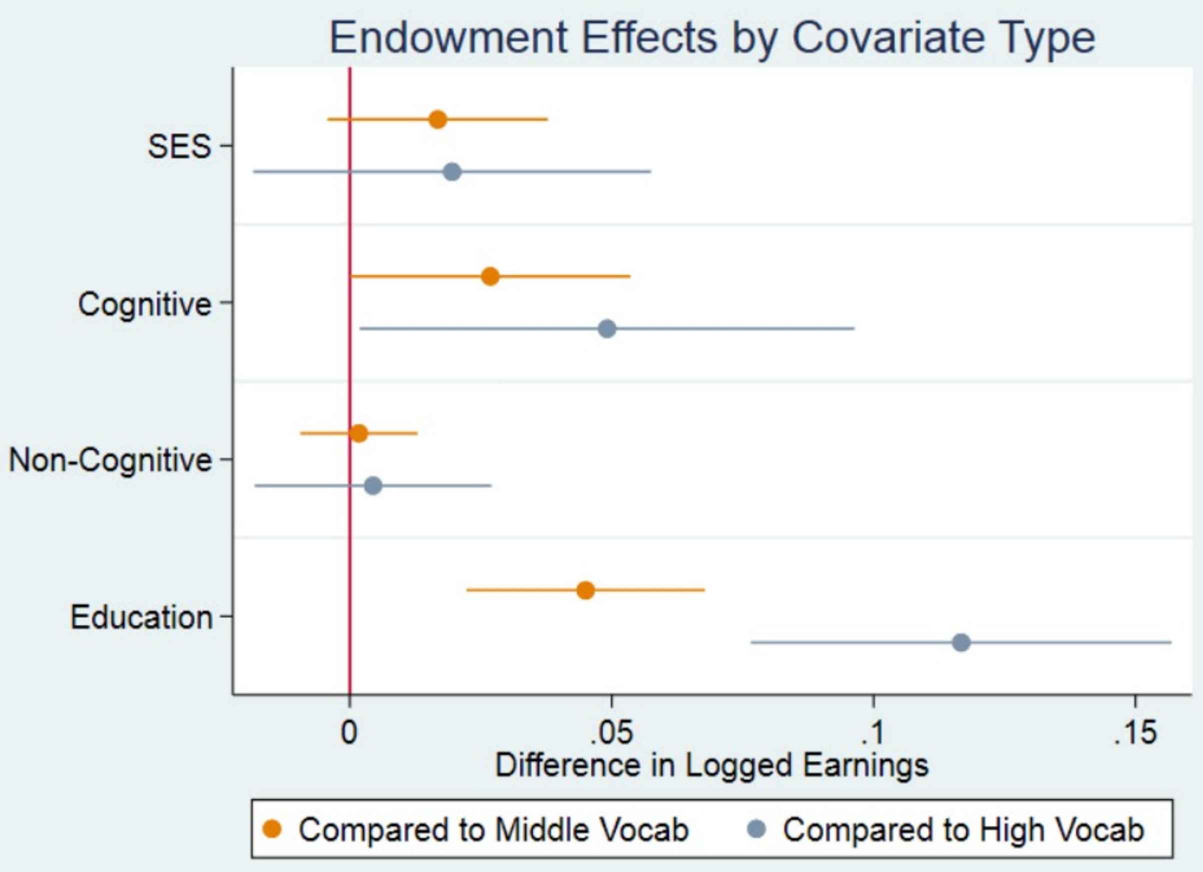


Figure 4-18 Decomposition of endowment effect by covariate type

Note: Vocabulary group cut-offs account for gender, so ratio of males to females is equivalent between groups. Therefore, no endowment effect of gender was anticipated, no effects were observed in Table 4.3 and we therefore do not plot them here. Plotted are the effects of variables which differ in mean levels between groups.

Whilst the endowment effect accounts for approximately two thirds (slightly less in the low to middle vocabulary comparison) of the total difference in earnings, the cumulative coefficient and interaction effects are jointly non-significant. That is, the visual point estimates in Figure 4.17 have a confidence interval which includes zero and we cannot reject the null hypothesis that vocabulary does not moderate the returns to other variables.

In general, the effects are similar but larger within the high vocabulary group although there is a notable large endowment effect of education when comparing these two groups. Analysis of specific covariates reveals that this is primarily driven by the substantially higher probability of obtaining a degree within the high vocabulary sample relative to both the low and middle vocabulary groups.

4.7.2 Gender differences

Figure 4.19 demonstrates the differential effects by gender. The main difference is the larger earnings gap for females, particularly in the high to low vocabulary group comparisons. This gap is primarily explained by the much higher endowments of the covariates for the high vocabulary females relative to the low vocabulary females.

In particular, Figure 4.20 shows that the largest portion of the low to high vocabulary earnings gap explained by the decomposition within the female sample comes from different endowments of educational attainment. Within this factor, there is the strongest effect observed for the probability of obtaining a degree. This particular covariate has a much stronger effect on the earnings gap within females than males.

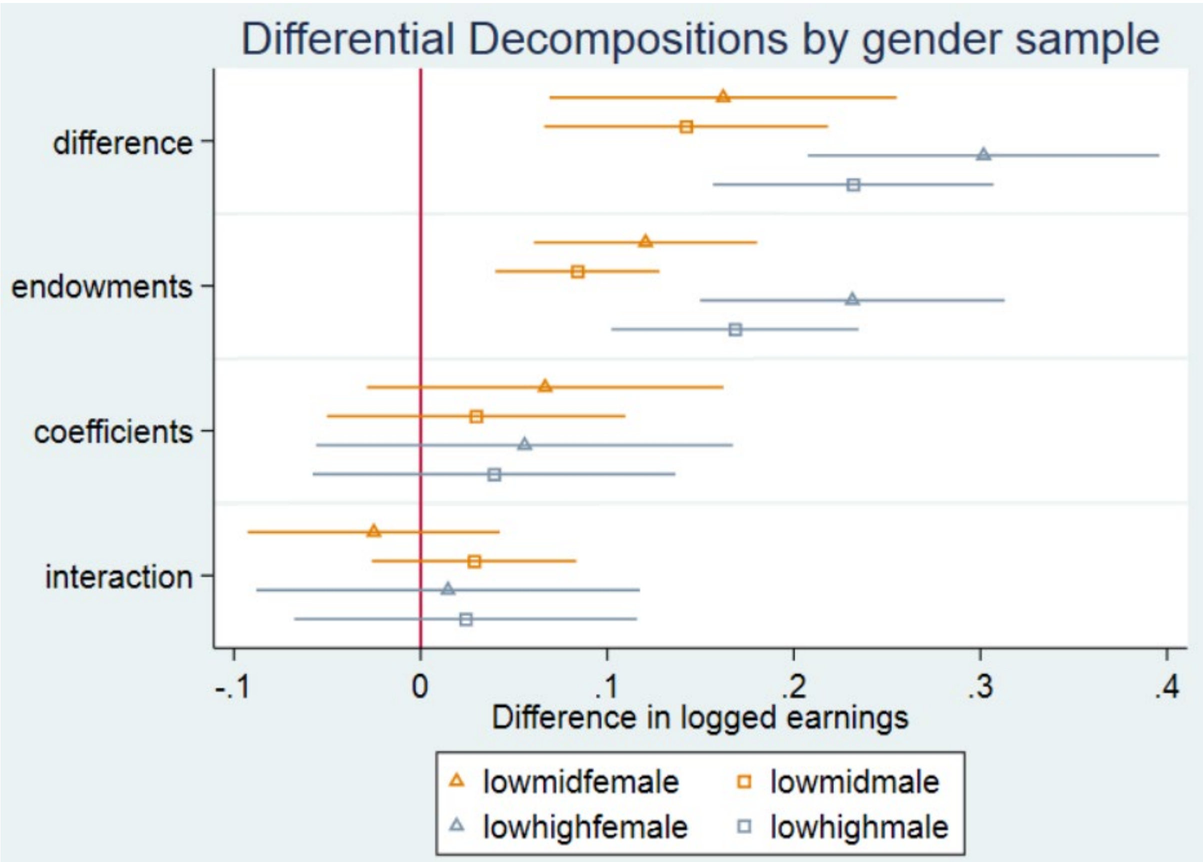


Figure 4-19 Male to Female sample comparisons of Oaxaca earnings gap decomposition into endowment, coefficient and interaction effects

Note: Overall threefold decompositions for female-only sample (triangle markers on top) and male-only sample (square markers beneath). Again, orange represents the middle vocabulary group comparison and blue represents the high vocabulary group comparison.

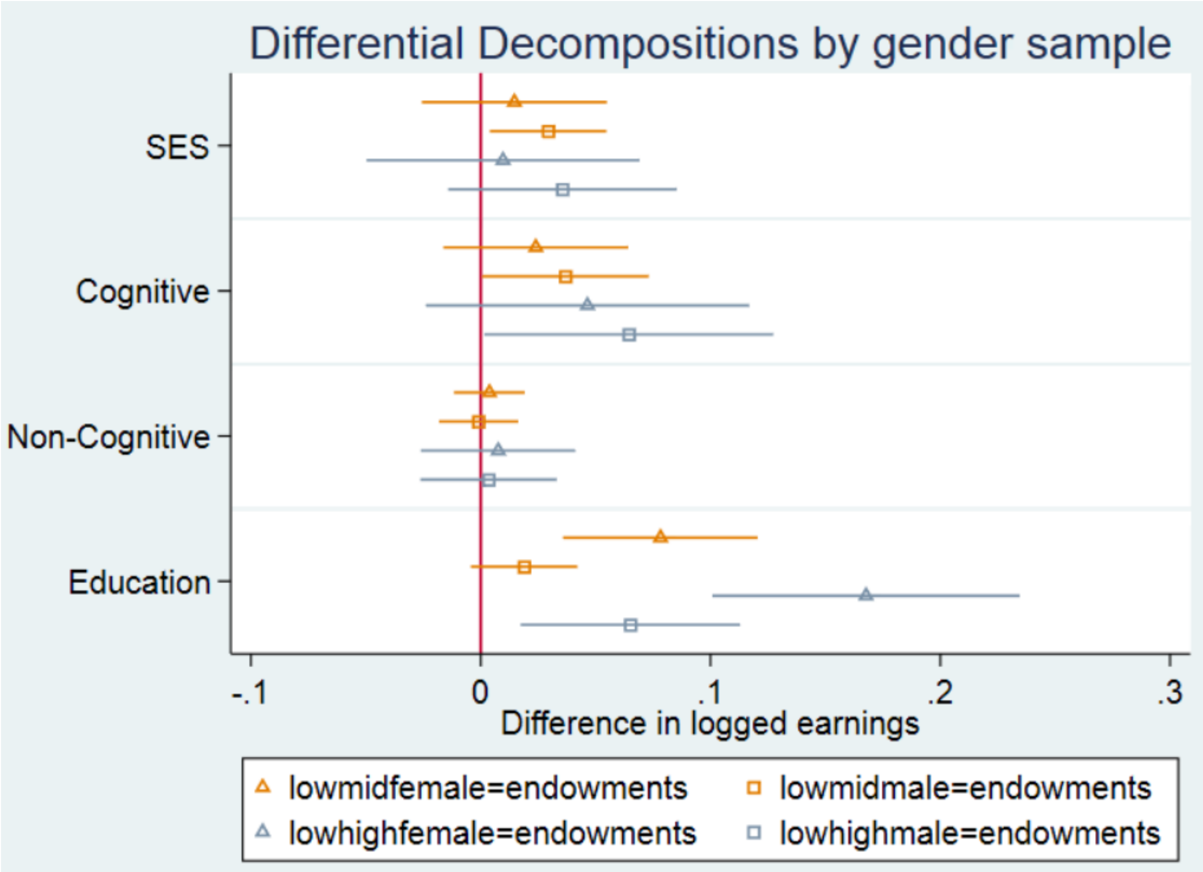


Figure 4-20 Male to Female sample comparisons of endowment effect separated by covariate type

Note: Orange represents a low to middle vocabulary group comparison and blue represents a low to high vocabulary group comparison. Triangle markers represent the female-only sample and square markers represent the male-only sample.

4.8 Discussion

Vocabulary clearly correlates to a range of important variables which predict later life success, particularly income in this application. In particular, this decomposition highlights that the correlation between income and cognitive ability through childhood as well as educational attainment helps to explain why there is a relationship between vocabulary and income.

The decomposition helps to estimate what a low vocabulary child would earn if they had equivalent covariates to a high vocabulary child. We found that the endowment effect of different mean levels of covariates explained 70% of the income gap. When comparing to a middle vocabulary child, the endowment effect explained 58% of the income gap. This suggests that vocabulary itself and the related unobservable factors add value to the understanding of the earnings gap observed. The covariates themselves do not explain the entire earnings gap, demonstrating that vocabulary is an important factor in predicting earnings alongside these covariates.

However, this decomposition finds limited evidence of the potential for early vocabulary to moderate the effects of other covariates. This is represented by the non-significant overall coefficient and interaction effects in each sample and comparison.

It is important to consider the effect of an intervention targeted at a child with poor receptive vocabulary. Improving a child's vocabulary is unlikely to occur in isolation. Human capital formation processes are endogenous and an improved vocabulary through intervention may well improve the child's reading and more general communicative ability. Therefore, it is not only the coefficient effect that we may observe by improving vocabulary. If intervention was able to improve a language impaired individual's educational outcomes, as hypothesised in Marsh et al.'s (2010) projected estimates of how intervention would impact on lifetime earnings, then intervention could increase earnings through the sizeable endowment effect related to education. This decomposition helps to highlight the importance of educational attainment as well as other cognitive and early academic performance measures related to early vocabulary.

Jones and Kelly (1984) summarise that the above makes a Oaxaca-Blinder decomposition a "mental experiment" rather than a "social experiment". Any policy aimed at effecting the returns to covariates, in our case to alter the vocabulary group of an individual, would be likely to cause spillover effects that impact the mean covariate levels in each group (Jones and Kelly, 1984). In our example, intervention would largely target vocabulary, but it is unlikely that the only impact would be observed within the coefficient effect portion of the earnings gap. Therefore, the coefficient effect is likely an underestimate of the change to earnings we would observe by increasing vocabulary sufficiently to make an individual move into a higher vocabulary group.

4.8.1 SES

The results in Figure 4.17 show that the differences in SES predict membership of the low vocabulary group but do not differentiate well between the middle and high vocabulary groups. This suggests that whilst low SES may be a predictor of low performance, high SES does not necessarily predict high performance. Therefore, the endowment effects of SES are equal in both comparisons and both are not statistically significant. The non-significant endowment effect of SES suggests that the benefits of high SES parents are largely controlled for by measures of cognitive and non-cognitive ability as well as educational attainment.

4.8.2 Cognitive

The results presented above demonstrate the value of vocabulary in predicting cognitive abilities which later predict income. Being in a high vocabulary group predicted higher mathematical ability at age 10 which was a significant predictor of income. However, the high multi-collinearity between maths and reading at age 10 makes the effect size relatively hard to estimate and instead we conclude that general cognitive human capital is strongly predicted by membership of a higher vocabulary group.

4.8.3 Non-Cognitive

Non-cognitive traits were relatively poorly predicted by membership of a high vocabulary group, relative to a lower vocabulary group. These same non-cognitive traits also poorly predict income within each sample. This suggests that age 10 may be too early to measure the non-cognitive traits which would be valued in the labour market in the long term.

Alternatively, it may reflect the specific non-cognitive skills available within the data, as there is very little research relating locus of control to language, and it is unknown at what age the early associations with conscientiousness (see Table 1.2) begin to disappear (see the results of Chapter 2).

Research within a sample containing well-responded and robustly measured non-cognitive skills would be beneficial to our understanding of the relationship between vocabulary and human capital.

4.8.4 Education

Our findings on performance at the end of compulsory education supports the literature on educational outcomes for children with poor vocabulary in childhood. Our results go further to suggest that subsequent differences in the probability of obtaining an undergraduate degree between low and high vocabulary groups are the strongest predictor of subsequent earnings. This supports the conclusions of Tomblin et al. (2008), that different socially valued outcomes between groups (education being a prime example) are the key indicator of the value of early intervention.

4.8.5 Gender

Obtaining a graduate degree was particularly important for females. However, the context of the data is significantly different from the current market for higher education which has expanded significantly since the early 1990s and had considerable growth in the proportion of students who are female. Therefore, our results may differ from those that could be estimated in more modern cohorts, although these would not have the benefit of multiple income observations across adulthood.

4.8.6 Method

The Oaxaca-Blinder decomposition is a useful tool for analysing differences between groups. However, it is subject to some sizeable caveats about its application and interpretation. In particular, the large influence of where the covariate is centred giving different coefficient effects is a major flaw. If a researcher wanted to find a significantly different return to a particular variable then this is relatively easy to achieve, sometimes referred to as “researcher degrees of freedom” (Wicherts et al., 2016). In our application, we make our intercept reference point as close to an individual with zero skills as possible. However, interpreting the intercept as a pure measure of group membership discrimination is problematic as no individual with zero for each measure exists within the data in any of the three vocabulary groups.

Further researcher decisions are reported in the appendix to demonstrate that our findings are robust. For example, the decision to split the vocabulary sample into thirds could alternatively have been split at the median, or quartiles or at -1 standard deviation and +1 standard deviation from the mean. Each of these specifications is reported in Appendix Figure A4.1 and our results remain broadly unchanged. The core result that the endowment effect accounts for the largest portion of the earnings gap, particularly cognitive and educational endowments and that the coefficient effect is positive but non-significant remain robust.

Another decision made was the reference group being the disadvantaged group, in line with the literature and intended interpretation. The alternative perspective is reported in Appendix Figure A4.2.

To conclude, the Oaxaca-Blinder decomposition demonstrates the importance of the relevant covariates which correlate to both early language skills and to lifetime earnings. We find that individuals with a higher level of vocabulary at age 5 experience a range of advantages. We also find that educational attainment from age 16 onwards is the most important target for individuals with poor or middle vocabulary skills aiming to close the earnings gap to high vocabulary individuals.

5 Chapter 5: The Use of Thresholds in Language Assessments

5.1 Introduction

Two important questions were posed in the first paragraph of this thesis. What does life look like for a child with early communication difficulties? How and why are their lives different from children with typical speech and language development? In the previous analyses, there have been two recurring themes; firstly, the importance of measurement in identifying the relevant population; and secondly, the breadth of impact over the life course.

In Norbury et al. (2016), the definition of the population is dependent upon the presence of a "functional impact". It is this impact which makes the population worthy of intervention. This analysis chapter uses a range of later life outcomes as a method of measuring the accuracy of a vocabulary test in correctly identifying the relevant population.

This builds upon the findings of Chapters 2 and 3 (differential outcomes for those with early communication difficulties) but is mostly motivated by the robustness check conducted in Chapter 4 which showed that the results of the analysis were not sensitive to altering the threshold lines drawn on a vocabulary assessment for defining high and low vocabulary. The use of thirds, quartiles, standard deviations, or other measures did not impact the core differences between groups, suggesting that there might not be an ideal threshold for detecting differences between groups.

Whilst the English Picture Vocabulary Test has been shown to predict outcomes in the previous chapters, it is unclear whether strictly-applied threshold values can accurately detect the presence of a functional language disorder at the margin. If it cannot, there are implications the way language assessments and threshold values should be used in research and practice. This chapter utilises a regression discontinuity design to address the following simple, widely applicable question:

Do threshold values on language assessments accurately identify language disorders? The answer to this question has a number of implications for how research and clinical practice are conducted. A language disorder is a disability which functionally impacts the child daily, explaining some portion of the wide range of associated difficulties in later life (Norbury et al., 2016; Bishop et al., 2017; Bercow, 2008). The impact of language disorders can be mitigated through effective intervention (Law et al., 2012b) and improving our understanding of these disorders through research (Law et al., 2017a). Accurate identification of language disorders is required in both settings. Children with language disorders are often not diagnosed and therefore do not receive the specialised assistance that could support them (Powell and Snow, 2012; Hendricks et al., 2019; Prelock et al., 2008).

Accurate diagnosis in both clinical and research settings is therefore a necessity. Currently, both clinicians and researchers utilise standardised, norm-referenced language assessments and relevant thresholds on a regular basis as at least one part of a diagnosis (Roulstone et al., 2015; Nitido and Plante, 2020).

The following literature review discusses the current measures for assessing the accuracy of language assessments and the application of threshold values for identifying language disorders. A regression discontinuity design is applied to the language assessment used in the previous two chapters as an empirical assessment of the validity of strictly applied thresholds on standardised language assessment scores for diagnosing language disorders.

5.2 Standardised language assessments and relevant thresholds

Standardised tests of vocabulary, both expressive and receptive, are frequently used by speech and language therapists in the UK and many other countries (Law et al., 2019a). Many norm-referenced standardised tests are explicitly targeted at non-specialist administrators, prioritising low cost and time input to enable scalability in resource-stretched clinical or educational settings. They can give a quick, easy, norm-referenced comparison for the performance of a child (Stockman, 2000; Brimer and Dunn, 1962; Dunn and Dunn, 2009).

Children's performance on standardised vocabulary tests have been consistently shown to strongly predict academic success, literacy, verbal intelligence and social mobility (Cunningham and Stavonich, 1997; Williams and Wang, 1997; Blanden, 2007). Standardised language assessments are useful research tools for indicating the language development of the child relative to their peers.

The RCSLT (Royal College of Speech and Language Therapists) recently surveyed all UK members for contributions in establishing the current UK research priorities for the field. The capability of screening tools for Developmental Language Disorder at various ages to accurately diagnose was felt to be largely unknown, with not enough research into the sensitivity and specificity of various tests. Whether language assessments can go further than providing a general ranking and accurately diagnose the presence of a language disorder is a fundamental question.

The ability of language assessments scores to provide diagnostic capacity rests upon a number of core assumptions:

1. The test must provide an accurate ranking of individuals.
2. Individuals either have the disorder or they don't.
3. A threshold must be present which separates individuals with the disorder (those who score below the threshold) from those without the disorder (those who score above the threshold) with some degree of accuracy.

The first assumption is questionable. Standardised test scores have predicted a range of outcomes in research, demonstrating that they provide some measure of underlying ability. Measurement error is present on any test and the score will proxy for the underlying ability on whatever facet of language the test is assessing (Jongsma, 1982; Bound et al., 2001).

According to McCauley and Strand (2009), a psychometric test should achieve a correlation coefficient of at least 0.9 on both interexaminer reliability (if two people test the same child, does the child score the same?) and test-retest reliability (if the same child sat the test at two different times of day, would the child score the same?). Cultural differences in the exposure to the words tested may also impact the way a vocabulary test ranks children, potentially leading to inaccurate rankings of underlying language ability (Washington and Craig, 1999). These are just a few ways in which a test may not always accurately rank individuals.

The second assumption is necessary for allocating service access ethically and is not the focus of this thesis or chapter. This assumption implies that a language disorder can be characterised as a ‘trait’, not a ‘dimension’, to use the psychology terminology (Pickles and Angold, 2003), although language assessments are commonly used as continuous measures of an underlying ability (Blanden, 2007; Almond et al., 2018). Similarities can be found in other health literature. For example, BMI scores may be used continuously or using threshold values to characterise risk (Bell et al., 2007; Pernar et al., 2016). This thesis has used both continuous and group identification measures at different times. Both modelling approaches have significantly predicted outcomes.

The third assumption is commonly tested for their accuracy within test manuals (Nitido and Plante, 2020; Plante and Vance, 1995; Letts et al., 2013), although these are generally not tested across the entire spectrum of scores (Spaulding et al., 2006). Thresholds at 1, 1.5 and 2 standard deviations below the mean are tested more than any others, intended to indicate mild, moderate and severe levels of disorder (Wiig et al., 1992), as detailed in Table 5.1.

Debate around the most appropriate threshold has been discussed for decades with multiple research papers using each threshold, with a minority also using a 1.25 standard deviation threshold relating to approximately the 10^{th} percentile. Tomblin et al. (1997) discuss this issue, noting the lack of consensus at the time, and little has changed in the subsequent decades (Law et al., 2017a).

Language assessment manuals may estimate the sensitivity and specificity of the assessment at a given threshold, usually at the common thresholds outlined in Table 5.1. Table 5.2 represents some fictional data used to demonstrate the calculation of four important values – sensitivity, specificity, positive predictive value and negative predictive value.

Table 5-1 Commonly applied thresholds or percentiles and associated levels of severity

Standard deviations below the mean	Approximate percentile (assuming a normal distribution)	Level of disorder (Wiig et al., 1992)
1	15.9	Mild
1.5	6.7	Moderate
2	2.3	Severe

Table 5-2 Hypothetical data example to illustrate key diagnostic terms

	No Language Disorder	Language Disorder	Total
Score Above Threshold	829	12	841
Score Below Threshold	101	58	159
Total	930	70	1000

Note: This is self-created data built exclusively for illustrative purposes. 1000 cases, of which 7% have a language disorder, are given a normal distribution rank by a test. A threshold drawn at 1 standard deviation below attempts to identify the presence of a language disorder. From this data, we could calculate sensitivity, specificity, positive predictive value and negative predicitive value.

Sensitivity represents the percentage of individuals who truly have the disorder who are correctly identified by the test's threshold.

$$\text{Sensitivity} = \frac{\text{N (With Disorder and a score below the threshold)}}{\text{N (With Disorder)}} = \frac{58}{70} = 82.9\%$$

Specificity represents the percentage of individuals who do not have the disorder who are correctly identified by the test's threshold.

$$\text{Specificity} = \frac{\text{N (Without Disorder and a score above the threshold)}}{\text{N (Without Disorder)}} = \frac{829}{930} = 89.1\%$$

Positive predictive value represents the percentage of individuals who the test identifies as 'disordered' who actually have the disorder.

$$\text{Positive predictive value} = \frac{\text{N (Score below the threshold + has disorder)}}{\text{N (Score below the threshold)}} = \frac{58}{159} = 36.5\%$$

Negative predictive value represents the percentage of individuals identified by the test as not having the disorder who actually do not have the disorder.

$$\text{Negative predictive value} = \frac{\text{N (Score above the threshold + no disorder)}}{\text{N (Score above the threshold)}} = \frac{829}{841} = 98.6\%$$

Plante and Vance (1995) recommend that a chosen threshold should maximise sensitivity and specificity, with a minimum value of 80% sensitivity and 80% specificity to be viable tools in practice.

If a more severe threshold is drawn, sensitivity would decrease and specificity would increase as a result of more people scoring above the threshold. To the author's knowledge, there are no general recommended guidelines for minimum levels of positive and negative predictive value.

Positive predictive value tells us the probability of the individual having a language disorder, given that the test says they have a language disorder. The risk of a false positive, or Type I error, is therefore one minus the positive predictive value. In this case, 63.5% of those identified as having a disorder would not actually have one.

Similarly, 1.4% of people (one minus the negative predictive value) identified as not having a disorder, actually have one, i.e. a false negative or a Type II error.

In this example data, the prevalence is lower than the threshold set. If prevalence is higher than the threshold, a higher rate of false negatives (underdiagnosis) and a lower rate of false positives (overdiagnosis) would be expected. Whilst papers sometimes recognise the change in the severity of the group identified when using a stricter threshold (Nitido and Plante, 2020), the relative economic costs of each of these mis-identifications is rarely discussed.

Spaulding et al. (2006) analysed the scores on a range of langauge assessments for a group of children with language disorders. Their scores fell -1.34 standard deviations (s.d.) below the mean on average across all assessments. However, most of the language disordered group scored within the normal range (within 1 s.d. of the mean) on most of the different language tests sat. In essence, some children with language disorders score well on some tests and some children without language disorders score badly. This suggests that the first assumption (tests accurately rank individuals) is not met by most of these tests. This drastically reduces the expected accuracy in diagnosis from any threshold drawn across the spectrum.

The effect observed in Spaulding et al. (2006), that children are hard to distinguish using the full spectrum of a given assessment, is likely to become amplified at narrow margins around a strictly applied threshold. The four values calculated from Table 5.2 which are commonly referred to in the literature do not account for the possible reduction in accuracy which occurs at the margin (near the threshold). If the accuracy is reduced at the threshold, this has ethical and empirical implications for how strictly a threshold can be used in a range of applications.

Threshold values are used primarily in three areas within speech and language sciences:

- Clinical diagnosis
- Research diagnosis
- Prevalence estimation

The following sections discuss each of these applications in turn.

5.2.1 Thresholds in Clinical Applications

A diagnosis of SLI or DLD commonly involves an informal assessment of the child by the therapist observing them at play, particularly for younger children, as well as a standardised assessment of their language ability with their performance compared to a norm-referenced sample using one of many tests (Roulstone et al., 2015; Letts et al., 2013; Bishop et al., 2016). The majority of referrals to speech and language therapy services come from teachers (Bishop et al., 2016), although parents, health visitors and other relevant adults can request the development of an Education and Health Care Plan for a child, including an assessment of learning difficulties.

Whilst a diagnosis is not entirely reliant on the score achieved by the child, assessment manuals are commonly used in clinical diagnoses which recommend the noted threshold values for identification. A score which matches an assessment diagnosis may then be used in securing funding for the services required by the child (Roulstone et al., 2015). For example, the New Reynell Developmental Language Scales (Letts et al., 2013) report the sensitivity and specificity of their assessment scales at 1.5 and 2 standard deviations below the mean within the manual for making a diagnosis.

Spaulding et al. (2006) found that, of 43 standardised assessment manuals reviewed, only 9 actually reported the sensitivity and specificity of the assessment at their recommended threshold value. Any diagnoses made using the recommended thresholds in the other 34 manuals would not be made with any measurable gauge of the accuracy of the diagnosis.

Anecdotally, clinicians have reported thresholds being applied at 2 standard deviations below the mean in clinical settings, a severe level of impairment (Wiig et al., 1992). This allows limited resources to be targeted at those most in need, assuming that the test scores are a good measure of the severity of the disorder. The relationship between available resources and the criteria for access to resources is likely to be endogenous: In areas of limited resource, fewer children are likely to be officially assessed and stricter thresholds may be applied. Together, this may dampen the estimated number of children who are in need of services in that local area (Strand and Lindsay, 2009). The lack of consensus on optimal threshold choices, the main choices of which are very far apart, is an important issue in addressing the postcode lottery associated with speech and language therapy in the UK (Longfield, 2019).

Law et al. (2019a) summarise the approaches to managing language disorders across Europe and beyond, finding that the use of strictly applied thresholds to language assessments is not limited to the UK. In Croatia, many SLPs were reported to use 1 standard deviation below the median score as identifier of impairment (Kraljević et al., 2019). In Ireland, language classes are only accessible to children with DLD who meet strict cut off points of 2 standard deviations below the mean on language assessments and also have a non-verbal IQ of above 90 (Frizelle et al., 2019). This criteria has largely been unchanged since 1993. This shows simultaneously the broad use of threshold values and the inconsistency in placement of the thresholds.

However, it is worth noting that even in situations where the guidance is to apply a threshold value strictly, 28% of therapists in Ireland stated that they did not apply this strict criterion to their clinical diagnoses (Frizelle et al., 2019). Roulstone et al. (2015) similarly report that some therapists are becoming less likely to conduct full assessments and apply threshold values, primarily due to time constraints within a first contact with the child. This may suggest that the impact of thresholds in clinical settings may decrease if this practice continues to become more common.

In a clinical context, false negatives are costly. Severe thresholds increase the number of false negatives within the population. Not providing support to a child with a language disorder limits the capacity to avoid a range of negative outcomes, as described in the literature are prior analysis chapters.

False positives, which are more common with milder thresholds, are also costly. There is a direct cost in removing a child from the classroom and providing services (Marsh et al., 2010), however the primary cost is an indirect opportunity cost. As services are stretched due to limited resources, provision of support to one child without a language disorder may mean that support is taken away from a child who does have a language disorder. For this reason, Bishop et al. (2016) do not recommend universal screening as the milder thresholds used tend to risk over-identification given that the prevalence of language disorders is typically lower than 16% (1s.d. below the mean, the mildest common threshold) at any given age.

Avoidance of these costs is a fundamental efficiency target. As manuals tend to test at the commonly applied thresholds only, this may miss the opportunity for mathematical optimisation of sensitivity, specificity, positive predictive value and negative predictive value, which may occur at different values along the distribution of scores.

5.2.2 Thresholds in Research

Within the literature reviewed in Section 1.2, many studies model group differences between those with language disorders and those without. Two methods are used to identify the groups. Some studies recruit the language disordered group from clinical settings who have received an official diagnosis (e.g. Conti-Ramsden et al., 2013). Other studies apply threshold values to language assessments and identify those below the threshold as having a language disorder (e.g. Forrest et al., 2018). The latter method implies the same assumptions outlined prior, that a threshold value on a language assessment can accurately identify a language disorder. These methods produce distinctly different samples – two thirds of children with SLCN (identified by test scores) do not present for help (i.e., request it in some way or get referred) (Tomblin et al., 1997; Law et al., 2013). This demonstrates the importance of 'functional impacts' for referral (Norbury et al., 2016).

When recruiting participants for intervention studies, threshold values can be applied as an exclusionary criterion in order to identify a target population, i.e., individuals with a language disorder. The 16th centile, or 1 standard deviation below the mean, is commonly used as an identifier. This relatively mild level of identification could reflect the difficulty in recruiting a sufficient sample size for many projects, whereas a clinician would have the opposite problem of oversubscribed services.

Children who score narrowly above a certain threshold would be characterised as a member of the group without language disorders in second-hand data research. These children would also be excluded from recruitment into intervention studies. If these thresholds correctly identify children at the margin (near the threshold) then this is not an issue. However, if a significant proportion of children are wrongly categorised near the threshold then this may impact the interpretation of the results of many research studies.

When a threshold with low sensitivity (low rate of true negatives) is applied, usually at a more severe threshold, those identified with the disorder by the test threshold may represent a more severely impacted group than the underlying sample with the disorder, leaving a biased sample. This stacks the deck towards finding significant differences between groups in research (Nitido and Plante, 2020).

The opposite can be said of thresholds with low specificity (low rate of true positives), typically milder thresholds used in screening procedures. Children without the disorder are more likely to be included in the disordered sample which makes it more difficult to find significant differences between groups (Nitido and Plante, 2020).

Threshold values are measured in terms of sensitivity and specificity when judging their diagnostic accuracy. These measures operate across the whole distribution of scores, not at the margin, but they do offer some indication of the reliability of the threshold value within the language assessment. Nitido and Plante (2020) found that, of 90 studies reviewed from 2015 to 2019, only 38 used a threshold on a language assessment with a known sensitivity and specificity. The majority of papers reviewed were using threshold values for language assessments that had not been validated in any form for that specific test. In summary, most recent research reviewed by Nitido and Plante (2020) identified groups as having language disorders without any statistical grounding for the grouping made. The fact that the results presented in Chapter 4 of this thesis are robust to changes in the threshold value show that this does not necessarily invalidate the findings of these papers, but Nitido and Plante (2020) state that researchers should align their measures with those validated in test manuals for clinical use.

5.2.3 Thresholds for Prevalence Estimations

Section 1.2.3 of this thesis summarises the methodologies for estimating the prevalence of Speech, Language and Communication Needs, or of language disorders specifically, with specific reference to the use of threshold values. Law et al. (2017, p17) report a range of prevalence estimates by age of children for the UK. Recognition is made of the impact that threshold cut-offs have on prevalence estimates, as in Law et al. (2000) and McKean et al. (2017b) – A stricter cut-off on a standardised test will naturally detect a more severe level of difficulty, implying a lower prevalence rate.

As outlined in Table 5.1, a cut-off of 1 standard deviation below the mean on one assessment will tend to result in a prevalence rate of approximately 16% and a cut-off of 2 standard deviations below the mean on a single assessment will tend to result in a prevalence estimate of approximately 2.3%, purely due to the characteristics of the normal distribution. For example, Randall et al. (1974) use a criterion of 2 standard deviations below the mean on the

Reynell Developmental Language Scales and find that 5 of 176 (2.5%) of children tested were "language impaired".

This endogeneity of using a distributional cut-off to estimate prevalence is problematic. The approach can be justified if the threshold chosen is shown to be accurate in detecting language disorders. However, as Norbury et al. (2016) report, prevalence estimates tend to be lower when requiring the child to score below a threshold on a number of tests. Whilst battery testing can help to identify areas of specific need, battery testing is unlikely to have been tested for sensitivity and specificity on relevant populations (Tomblin et al., 1997 as an exception).

When comparing prevalence rates, it is important to be aware of the influence of threshold placement on the observed prevalence. For example, in the Netherlands, 8 studies found prevalence rates ranging from 2.8 to 27 percent for DLD in children aged 0 to 7 years (Gerrits et al., 2019). One of the reasons for this wide range was the use of different thresholds on the language screening instrument.

A lack of standardised assessments in Finland has been linked to low observed prevalence rates (< 1%) (Kunnari et al., 2019). Language assessments may pick up underlying difficulties which do not present in an obvious manner. There are no reported examples of thresholds on language assessments being drawn below the first percentile. In this sense, language assessments and thresholds may help to avoid situations of systematic underdiagnosis.

For a threshold value on a language test to accurately estimate prevalence, the test must have broadly equal sensitivity and positive predictive value. If this is the case, the number of children with language disorders wrongly identified as not having them will broadly equal the number of children without language disorders wrongly identified as having them. Therefore, the prevalence estimate will be relatively accurate, regardless of whether the test identified the correct children. However, to equalise these requires a threshold placed exactly at the prevalence rate. This is either unknown at the outset, hence the need for estimation, or is a sample prevalence rather than a population level prevalence. For example, Letts et al. (2013) use a community sample to test the identification accuracy of the New Reynell Developmental Language Scales. A threshold which equalises sensitivity and positive predictive value within this sample would be set at the prevalence rate within this sample, not the whole population.

Bishop et al. (2016) state that threshold values from screening procedures tend to substantially over-estimate the prevalence rate due to a relatively weak positive predictive value (many of those identified with the disorder do not actually have the disorder). This is because screening procedures, as a first stage in a diagnostic process, tend to utilise mild thresholds as they are primarily concerned with avoiding false negatives rather than false positives, as false positives can be identified correctly at a later stage in the diagnostic process.

5.2.4 Thresholds at the Margin

There are many issues noted in the literature relating to the use of thresholds and their accuracy across the distribution (sensitivity and specificity). Evidence-based practice usually relies upon these figures and the recommendations in Plante and Vance (1995) to justify the use of a strict threshold. Accurately recognising that a child who scored 100% does not have a language disorder is included within the measure of the accuracy of the test.

Strict thresholds are justified through high sensitivity and specificity. However, sensitivity and specificity are likely to be reduced when analysing a narrower range of scores. Where thresholds are applied indiscriminately, grouping children scoring narrowly above the threshold equally to those scoring far above the threshold and vice versa, the likelihood of misdiagnosis is not represented at the margin. Therefore, whether language assessments can accurately diagnose language disorders at the margin (near the threshold) has not been tested. The robustness of research findings to narrow changes in the applied thresholds are not typically reported, limiting our knowledge of the value of thresholds at the margin. The findings in Chapter 4 suggest that narrow changes in the (non-clinical) threshold do not impact findings, implying that children near to a threshold may be poorly differentiated from one another.

Language disorders are characterised as representing a functional difference in a child's capabilities (Le et al., 2020b; Great Ormond Street Hospital, 2020; Bishop et al., 2017). Therefore, if a strictly applied threshold could identify language disorders, there would be different expected outcomes for children either side of the threshold. If the threshold could accurately identify at the margin, i.e. near the threshold, this would represent a sudden increase in expected outcomes when crossing over the threshold score. If this is not the case, there is little justification for applying different procedures and policies to children either side of a threshold at a narrow margin. This analysis formally tests for differences in outcomes between individuals at the margin in the absence of corrective intervention.

5.3 Research Questions

5.3.1 Stage 1

1. Can the most commonly applied threshold for the English Picture Vocabulary Test accurately identify functional difficulties requiring intervention?

5.3.2 Stage 2

2. Can any alternative threshold applicable to the English Picture Vocabulary Test accurately identify functional difficulties requiring intervention?

5.4 Data

5.4.1 British Cohort Study 1970

The data is collected from the sample sample as that used in Chapters 3 and 4. The British Cohort Study is a UK cohort of 17,827 children (9841 adults remaining) born in one week in April 1970 and followed up at ages 5, 10, 16, 26, 30, 34, 38 & 42.

Information was collected in childhood on many socio-economic characteristics of the family, early information on their development and then through into academic attainment, occupational success and health outcomes.

5.4.2 English Picture Vocabulary Test

The same English Picture Vocabulary Test scores at age five from the previous two analyses are used again, this time with a threshold value applied to attempt to identify language disorders. Age five is considered old enough for a reliable diagnosis from a language assessment (Letts et al., 2013).

Scores are distributed as either a 0 or any integer from 5-56 (see Section 3.2.3 for a more detailed description of the scoring process). The distribution of scores is shown in Figure 5.1, with the strict upper ceiling of 56 questions explaining the short right hand tail.

The same exclusion criteria used within Chapters 3 and 4 apply, that only children with English as the first and home language are included as per the instructions in the original test manual and the modern manual (Brimer & Dunn, 1962; Dunn & Dunn, 2009). Only children tested within three months of their fifth birthday were included, as in Chapters 3 and 4, to enable comparability. The results presented are robust to the removal of each of these restrictions.

The Picture Vocabulary Test is commonly used in the UK (Roulstone et al., 2015) for current identification of language disorders and multiple versions exist in other languages.

Jongsma (1982) show the median measurement error on the original American version of this test, the Peabody Picture Vocabulary Test, to be 7 standard points, or around half a standard deviation, demonstrating why strict thresholds may be an ineffective diagnostic technique.

5.4.2.1 Threshold chosen for this analysis

The test used within the British Cohort Study is an amalgamation of the original English Picture Vocabulary Test aimed at school-age children (from five upwards) and that aimed at pre-school children (up to age five). Therefore, it has no recommended diagnostic threshold. However, as Nitido and Plante (2020) report, many studies still apply strict thresholds to tests without a measure of the diagnostic capabilities of their chosen measure and suggest an accurate identification of a group with language disorders.

One standard deviation below the mean is the most commonly applied threshold to the English Picture Vocabulary Test or Peabody Picture Vocabulary Test (Washington & Craig, 1992; Washington & Craig, 1999; Law et al., 2009) and so this is the threshold we apply for the Stage 1 analysis. This has an added benefit of giving a larger sample size for comparisons around the threshold, as fewer individuals group near the threshold at more extreme values.

As shown in Figure 5.1 and Table 5.3, all values of 24 and below are characterised as having the language disorder. All values of 25 and above are characterised as not having the language disorder.

Stage 2 extends this to test the other common thresholds as well as other arbitrary thresholds across the whole profile of results.

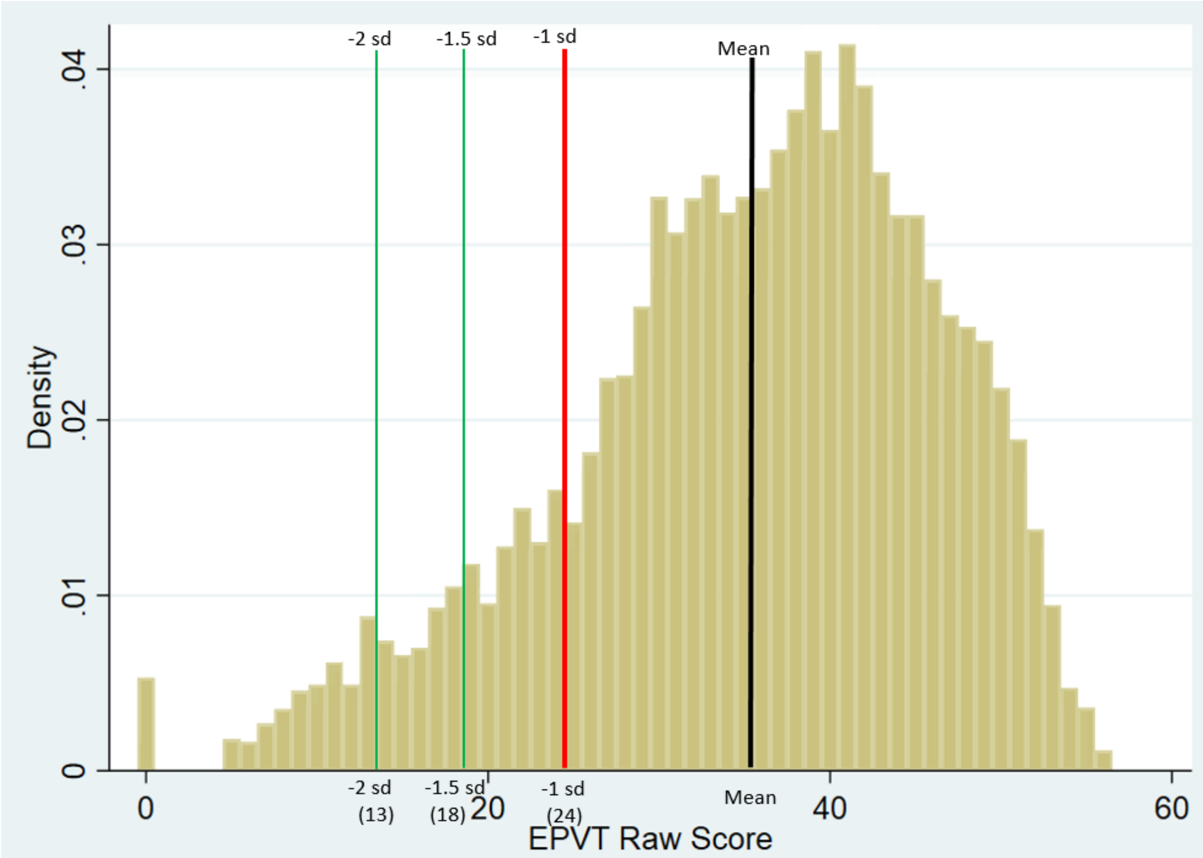


Figure 5-1 EPVT Distribution with means and common threshold values plotted

Note: Standard deviation of the full sample used to calculate thresholds.

Observations	Mean	Standard Deviation	Minimum	Maximum	-1 s.d.	-1.5 s.d.	-2 s.d.
12,289	35.2	10.8	0	56	24.4	18.9	13.6

5.4.3 Outcome

The functional impact that a language disorder has on an individual may relate to a number of aspects of life. Johnson and Kussbykh (2008) reference the potential impact of early child language on broad 'life chances', determining the overall quality of life experienced by an individual. Therefore, a wide range of outcome variables are taken at different ages to assess whether the threshold value applied accurately detects differences between individuals above and below the threshold. The following types of outcomes are considered across the life course:

- Language
- Education
- Occupational
- Quality of Life / Health

The full range of outcome measures tested within this analysis, intended to capture a broad range of potential impacts of a language disorder, are summarised within Table 5.3.

Chapter 1 reviews the literature relating to each of these spheres of life, detailing why early communication difficulties (in this case, a language disorder) may have impact over the life course.

5.4.3.1 Language outcomes

A language disorder, in the absence of specific intervention, is likely to impact upon later language development. If this were not the case, then intervention would not be necessary for these individuals. Vocabulary development may be impacted by initial vocabulary (Taylor et al., 2013; Reilly et al., 2010; McKean et al., 2017b), potentially additionally through factors such as enjoyment of reading (Snow, 2020).

The verbal similarities sub-scale of the British Ability Scales (Elliott et al., 1997) is selected as a measure of verbal ability at age 10. In this test, 20 sets of three related words (e.g. orange, banana, strawberry) are read aloud to the child and the child is asked to express a word consistent with the theme (e.g. apple) (Parsons, 2014).

A Vocabulary Test, created for the British Cohort Study, is selected as a measure of vocabulary at age 16. In this test, the child is given an initial word (e.g. reward) and five options (e.g.

notice, golden, prize, stable, marine), from which they have to select the correct synonym (Parsons, 2014). There are 75 initial words of increasing difficulty for the child to attempt.

The same vocabulary test was repeated at age 42 using a shortened version with only 20 initial words.

Collectively, these three measures model language development. Age 10 and 16 are the most temporally proximal waves available after the initial English Picture Vocabulary Test score at age five. Age 42 is the most temporally distal available vocabulary measure.

5.4.3.2 Educational outcomes

A language disorder, in the absence of specific intervention, may limit an individuals capacity to succeed educationally (Norbury et al., 2016).

Educational success at age 16, after the completion of compulsory education, was measured in three ways. Firstly, the number of o-levels or CSEs achieved by the individual. Secondly, the achievement of an o-level or CSE in English Language. Thirdly, the achievement of an o-level or CSE in Mathematics. These remain an important barometer of success in compulsory education today.

Educational success across the life course was measured using the highest academic qualification achieved. As a regression discontinuity design is a linear technique, conditional probabilities of achieving key educational milestones are modelled, as opposed to an ordered logistic model. Therefore, the highest academic qualification is again decomposed into the achievement of CASMIN Level 2 (5 or more CSEs or O-Levels) as a binary indicator and achievement of CASMIN Level 3 (a degree) as a binary indicator.

5.4.3.3 Occupational outcomes

A language disorder, in the absence of specific intervention, may limit an individuals capacity to succeed in the labour market (Clegg et al., 1999; Law et al., 2009).

Occupational success was measured in terms of the earnings of the individual (measured using the same midpoint estimation as in Chapters 3 and 4) as well as the occupational social class associated with the job.

Earnings were measured using the average of the earnings reported at age 30, 34, 38 and 42 for those who respond to all waves. Earnings were also measured at age 30, as the most well-responded timepoint. Earnings at age 30 and 34 were averaged, as the two highest responded timepoints. Across these three measures, we obtain a range of sample sizes and levels of accuracy to best reflect the earnings potential of the individual.

Occupational class is commonly measured as a categorical variable, either through National Statistics Socioeconomic Class (NS-SEC) or Registrar General Social Class measures. As the regression discontinuity design is dependent upon a linear model, the social class of the individual is dichotomised. The registrar general occupational social class of the individual at age 34 and 42 is dichotomised into a binary indicator of blue collar (unskilled, partly skilled and skilled manual) and white collar (skilled non-manual, managerial/technical and professional) occupations. Those who work in a white collar occupation at either time point are categorised as white collar, a representation of the highest occupation category achieved in adulthood. The linear regression then predicts the conditional probability of the individual working in a white collar occupation.

These four measures cover the breadth of income observed as well as proxying for the employment conditions of the individual across a 12 year period of adulthood.

5.4.3.4 Quality of Life Outcomes including Health

A language disorder, in the absence of specific intervention, may impact the quality of life an individual experiences into adulthood (Clegg et al., 1999).

Life satisfaction is measured on a 0-10 scale at age 34, simply capturing a proxy of the happiness and individual feels with their life’s opportunities and outcomes.

The GHQ-12 is used as a measure of psychiatric risk, capturing an element of overall psychological well-being at age 34. This is a shortened version of the GHQ-28 scale used within Chapter 2.

Mental well-being is captured using the Warwick Edinburgh Mental Well-Being Scale at age 42, a series of 14 questions on 1-5 likert scale responses.

These three measures capture an individual’s subjective well-being into adulthood.

Table 5-3 Outcome variables covering four areas of functional importance

Note: All outcomes are scaled cohesively, that a higher score represents a positive outcome.

Variable (Age measured)	Scale
Language	
British Ability Scales: Verbal Similarities Scale (10)	0-20
Edinburgh Reading Test Vocabulary Score (16)	0-75
Shortened Edinburgh Reading Test Vocabulary Score (42)	0-20
Educational	
Number of O-Levels	0-15
English Language CSE/O-Level (16)	0-1
Maths CSE/O-Level (16)	0-1
CASMIN Level 2	0-1
CASMIN Level 3	0-1
Occupational	
Log Annual Income (30)	(2.48 – 14.75)
Time-Average Log Annual Income (30 & 34)	(5.61 – 12.68)
Time-Averaged Log Annual Income (30 & 34 & 38 & 42)	(7.83 – 11.94)
White Collar Occupational Class [Registrar General] (34 & 42)	0-1 (V / IV / III-M) – (III-NM / II / I)

Variable (Age measured)	Scale
Quality of Life / Health	
Life Satisfaction (34)	0-10
GHQ Total Psychological Wellbeing (34)	12-45 [High score is better health]
Warwick Edinburgh Mental Well-Being Scale (42)	14-70 [High score is higher well-being]

5.5 Methodology

5.5.1 Stage 1

A Regression Discontinuity Design (RDD) tests the existence of a jump in the linear relationship between two associated variables at a given point. This is an "ex-post facto experiment". Individuals are not randomly assigned to a treatment but an empirical method is used to create treatment and control groups that are otherwise comparable at baseline (Villamizar-Villegas et al., 2020). Lee and Lemieux (2010) describe an RDD as the "close cousin" of a randomised control trial.

In this application, the treatment effect is the identification of a language disorder using a threshold value on the EPVT. The control group are those who score above the threshold. By comparing those very close to the threshold, the treatment and control groups are comparable on all observable covariates.

RDDs are used in the field of Education more commonly than any other, although there is no prior application within the field of speech and language sciences (Villamizar-Villegas et al., 2020). Two types of RDD exist, fuzzy and sharp:

In a fuzzy RDD, the threshold indicates a change in the probability of treatment, yet participants either side of the threshold may be in either group. This may be the true scenario of the impact of thresholds in clinical applications relating to language disorders.

In a sharp RDD, the threshold indicates a strict eligibility criteria. All individuals one side of the threshold, in this case below it, are in the treatment group, and all individuals above the threshold make up the control group.

For this hypothetical experiment, we use a sharp RDD, as we do not observe the actual presence or absence of a language disorder. All individuals below the threshold are identified by the EPVT as hypothetically having a language disorder. All individuals above the threshold are identified by the EPVT as not having a language disorder. The threshold perfectly, and deterministically, creates a treatment group:

Treatment = Threshold Identified Language Disorder (TI-LD group) = Scores of 24 and below.

Control = Threshold-Identified No Language Disorder (TI-NLD group) = Scores of 25 and above.

Close to the threshold, a natural experiment occurs. The group immediately above the threshold act as a control group for the group immediately below the threshold. Through this methodology, the TI-NLD group outcomes represent what would occur for the TI-LD group if they did not have a language disorder. If the two groups have significantly different outcomes, then the threshold is accurately identifying (at least a higher probability of) a language disorder which has a functional impact on the individuals in that group.

5.5.1.1 Self-Selection

The most common difficulty in sharp RDD applications is self-selection. Where eligibility changes at a threshold, individuals may be incentivised to intentionally score in a way which benefits them. For an RDD to be applicable, agents must be unable to precisely control the running variable around the threshold to enable randomised variation (Lee and Card, 2008).

Within this application, children sat the EPVT with no incentive to score one side of the threshold rather than another. Scores were not used to determine any access to services and the standardisation used for identification relates only to the sample analysed in this current application. Self-selection is therefore not an issue in the current application. This is demonstrable by the absence of any significant clustering of scores within Figure 5.1, as recommended by McCrary (2008).

5.5.1.2 Randomised Assignment

In order to estimate the causal impact of the threshold, individuals close to the threshold must be comparable at baseline. Children with scores near the threshold are therefore t-tested for group differences on a range of observable covariates. The group of participants who score between 22-24 and the group of participants who score between 25-27 are found to have broadly equal levels of parental and paternal social class, gender (distribution of male:female), parental education, probability of being a firstborn, geographical region distributions, probability of having seen an SLT, and average gestation period (full t-test results shown in Appendix A5.1). By testing on this wide range of observable covariates, we assume that all unobservable covariates are equal between groups. This assumption underpins the causal inference of the Regression Discontinuity Design. Those above the threshold then act as counter-factuals for those below the threshold and their outcomes show us what would happen to an individual with the disorder if they didn’t have it.

The only observable difference between those scoring 24 and those scoring 25 is their score on the EPVT (in line with the linear trend) and the identification of a language disorder at the 24 and below threshold. The RDD explicitly compares those scoring 24 to those scoring 25, testing for a discontinuous jump in the linear relationship as causally explained by the presence of a language disorder. The comparability of covariates between those scoring 22 and those scoring 27 suggests that using each score as an individual "bin" (Lee and Lemieux, 2010) is a robust quasi-randomisation procedure comparing individuals who are equivalent at baseline aside from treatment.

5.5.1.3 Linear modelling and inclusion selection

A linear relationship is estimated between the continuous running (assignment) variable, the EPVT score, and the outcome (see Table 5.3 for the list of outcome measures tested).

Whilst there are situations where other functional forms, including polynomials, have been used in a small number of applications, a linear estimation is suitable in the circumstances where only a narrow bandwidth around the threshold on the running variable are included (Lee and Lemieux, 2010). This is a trade-off, as by only including those at the exact threshold, the data is limited and the model simply becomes a t-test of two groups. As we have a large sample and a range of outcomes, a narrow inclusion bandwidth is preferred at five scores either side of the threshold.

The impact of the threshold treatment for a discontinuous 'jump' in outcomes is formally tested by Tau, our parameter of interest.

There is an additional interaction term between the treatment term and the running variable, testing for changes in the relationship between the running variable and the outcome associated with the treatment, as recommended by Lee and Lemieux (2010). This is included to reduce the number of assumptions associated with the design. We are testing whether the conditional expectation of the outcome changes when approaching the threshold from below, or above, and do not need to constrain that the gradients of both approaches are equal.

5.5.2 Model Specification

$$Y_i = \alpha + \tau D_i + \beta_1 X_i + \beta_2 X_i D_i + \varepsilon_i$$

$$D_i = 1 \; if \; X_i \leq \bar{c}$$

$$D_i = 0 \; if \; X_i > \bar{c}$$

$$\bar{c} = Threshold\ value$$

$$X_i = EPVT\ raw\ score$$

$$D_i = Threshold\ Identified\ Language\ Disorder$$

$$Y_i = Outcome\ score\ (15\ outcomes\ tested\ separately)$$

$$\alpha = Regression\ Intercept$$

$$\tau = Additional\ Intercept\ below\ threshold\ (Treatment\ Effect)$$

$$\beta_1 = Regression\ Gradient$$

$$\beta_2 = Additional\ Gradient\ below\ threshold$$

Within Stage 1, $\bar{c}$ is equal to 24, at one standard deviation below the mean. In Stage 2, the analysis is extended to include values of $\bar{c}$ at all integers from 9 to 35.

This model is repeated across a wide range of outcome measures to test the threshold's ability to predict differences in ability and outcomes. Outcomes are all coded positively such that a negative Tau would represent negative outcomes for the impaired group.

Only individuals who score within five of the threshold are included in the Stage 1 analysis. This has two major benefits:

1. Covariates do not need to be included in the regression, as they are equal between groups from scores of 22 to 27 (see Appendix Table A5.1) and therefore do not bias our treatment effect at the threshold.

Hahn et al. (2001) state that covariates which are not evolving smoothly with respect to the threshold must be included in the regression in order to avoid a Type I false positive error, where the treatment effect is biased by the omission of covariates.

2. We can assume a linear relationship between the running variable and the outcome.

If the relationship between the conditional mean of the outcome and running variable is not linear, a linear specification will be suitable over a narrow bandwidth, although an increasingly bad approximation may occur over a large inclusion bandwidth (Lee and Lemieux, 2010).

There is a trade-off between a narrow bandwidth, which reduces the potential impact of non-linearities, and a wider bandwidth, which increases sample size. There is no formal calculation for selecting a bandwidth and this selection was made pragmatically to balance sample size with the risk of non-linearities. Alternative bandwidths are tested to formally test the robustness of results.

5.5.3 Stage 2

Stage 1 tests whether a threshold at 24 or below can identify functional differences in language. Differences in outcomes for children in the control group (TI-NLD) and the treatment group (TI-LD) are used to indicate functional differences identified by the threshold. 15 outcomes are tested to cover a wide range of possible functional impacts. This threshold is the most commonly applied threshold to the test in question.

Stage 2 seeks to identify whether the results of Stage 1 are applicable at other thresholds. Commonly, in RDD applications, authors will test the presence of "placebo" discontinuities when the running variable crosses another non-relevant threshold (Imbens, 2004). This is to test the uniqueness of the applied threshold relative to any other.

The current analysis extends this principle further by testing the presence of discontinuities at all feasible thresholds. This method has not been used previously, to the author's knowledge. This innovation is recommended in future for avoidance of the potential for researchers to select placebo discontinuities which would support the main findings.

The model specification applied in Stage 1 is equally applicable in Stage 2. As stated in the model specification, only the value of $\bar{c}$ changes.

5.5.3.1 The range of thresholds tested

The lowest value of $\bar{c}$ tested is at 9. This is the lowest threshold testable which would include the minimum non-zero scorers as the same inclusion bandwidth is maintained.

The highest value of $\bar{c}$ tested is at 35. This is the highest feasible threshold, as the language disorder group here represent anyone below the mean. Therefore, it was not considered necessary to test any higher thresholds as this would be identifying language disorders for children performing above the mean which would not be relevant to questions of clinical disorders.

All integer values between 9 and 35 on the EPVT distribution are tested as potentially viable thresholds.

5.5.3.2 The bandwidths for inclusion tested

In the Stage 2 application, the trade-off between sample size and the risk of non-linearities is reassessed. At the low end of the distribution, a small inclusion bandwidth may result in a very low sample size. However, a larger inclusion bandwidth reduces the number of thresholds that can be tested at more extreme levels than the common 2s.d. threshold drawn from the literature. A bandwidth of five either side of the threshold is maintained, hence the lowest threshold value of 9.

The analysis is tested with inclusion bandwidths of three to seven either side of the threshold as a robustness check.

5.5.3.3 Type I Error

With 15 outcomes and 26 potential thresholds, we anticipate significant results purely through Type I error. Whilst these could be corrected for using Bonferroni standard errors (multiply the p value by the total number of regressions), such a large number multiplication would dramatically increase the chance of a Type II error (observing no significant results where true effects exist). This would limit our potential to compare different thresholds. The trade-off between Type I and Type II errors is directly impacted by the number of regressions ran and Armstrong (2014) do not recommend using Bonferroni corrections for each analysis.

As a solution, Bonferroni corrections are not applied to each regression. All significant results at any given threshold are viewed in the context of the probability risk of Type I error. This allows comparison of thresholds in their predictive capability. When a universal null hypothesis that all results are not significant is tested, a Bonferroni correction is applied.

5.5.3.4 Robustness to fluctuating covariates

Gender assigned at birth and the probability of visiting an SLT in childhood both fluctuate randomly across the EPVT distribution. Therefore, at some thresholds, the covariates may not be statistically equal at baseline across treatment groups and this could impact results. As a robustness check, the analysis carried out on the full sample was repeated within eight restricted samples – male only, female only, SLT-visited only, no SLT-visited only, male+SLT, female+SLT, male+no SLT, female + no SLT – to ensure that the results are not driven by covariate distributions.

5.5.4 Hypotheses

5.5.4.1 Stage 1 Hypothesis

Null Hypothesis H_0: $\tau_{24,\,1\text{-}15} \geq 0$

The one standard deviation threshold has no predictive power over outcomes 1-15.

Alternative Hypothesis H_A: $\tau_{24,\,1\text{-}15} < 0$

The one standard deviation threshold is an accurate predictor of difficulties for at least 1 of the 15 outcomes.

If τ is negative, this suggests that those identified as having a language disorder (TI-LD), i.e. those below the threshold, have significantly worse outcomes than those without a language disorder (TI-NLD), i.e. those above the threshold.

If τ is positive, this implies that the threshold identifies group differences but that it is not identifying functional language difficulties, as this is not the expected direction. Therefore, a significant positive is considered equal to a non-result as it does not support an affirmative response to the research question.

5.5.4.2 Stage 2 Hypothesis

Null Hypothesis H_0: $\tau_{c,\,1\text{-}15} \geq 0$

No value of $\bar{c}$, the threshold, has predictive power over outcomes 1-15.

Alternative Hypothesis H_A: $\tau_{c,\,1\text{-}15} \neq 0$

At least one threshold value, $\bar{c}$, is an accurate predictor of at least one of the 15 outcomes.

A Bonferroni correction will be applied to the testing of this universal null hypothesis to account for the large number of regressions tested.

There are four possible conclusions to the results of Stages 1 and 2, outlined in Table 5.4.

Table 5-4 Summary of potential conclusions from two stage analysis

Note: Combined conclusions of both stages' results of the accuracy of strictly applied thresholds for identifying language disorders at the margin using the BCS70 and the EPVT.

		Stage 1 Research Question	
		1 s.d. is effective	**1 s.d. is ineffective**
Stage 2 Research Question	**At least one other threshold is effective**	Strictly applied thresholds can identify language disorders, but a choice is required.	Strictly applied thresholds can identify language disorders. The wrong threshold was previously used.
	All other thresholds are ineffective	A one standard deviation threshold uniquely identifies language disorders.	No strictly applied threshold identifies language disorders.

5.6 Plotting the Discontinuity

A true discontinuity is commonly visible from plotting average outcomes by the variable in question, with no other similarly visible discontinuities. See Figures 5.2, 5.3 and Appendix Figure A5.4 & Appendix Figure A5.5 for examples of this with our data.[4]

We separate the participants into groups based on their score with each possible score having their own group of people who scored the same. By generating the mean outcome value for each group, we can then produce a scatter plot with a local polynomial plotted before and after the threshold we apply to look for evidence of discontinuous jumps in outcomes.

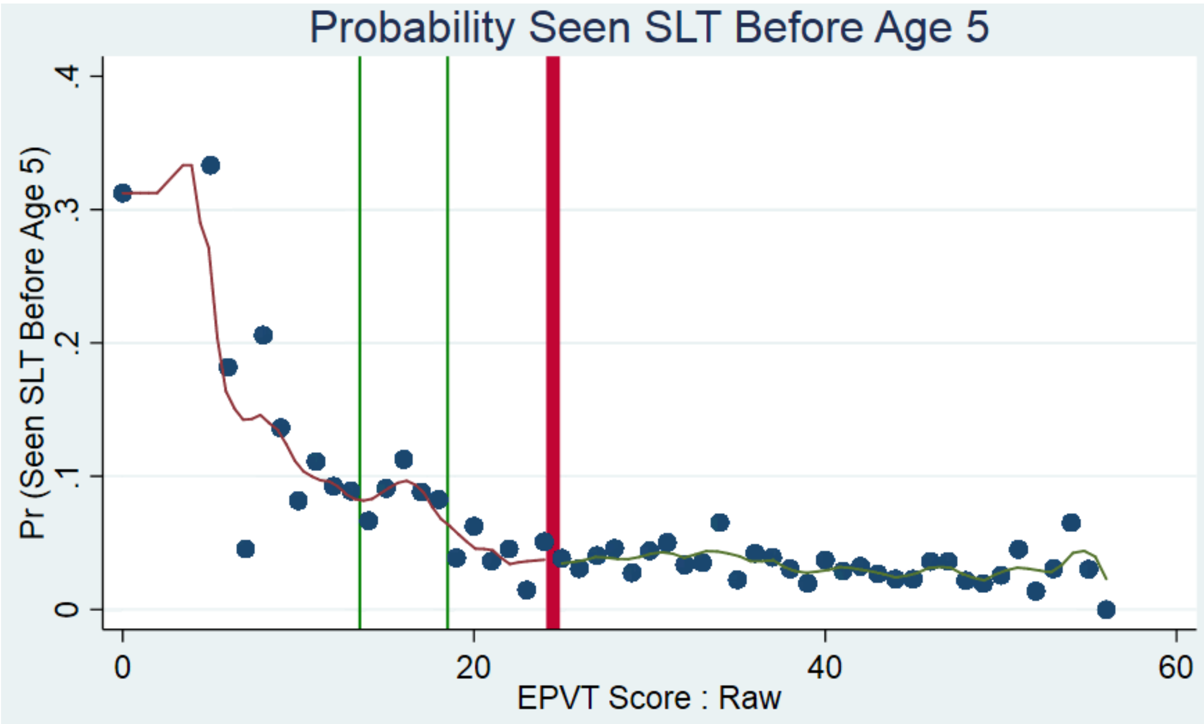


Figure 5-2 Local polynomial of probability of seeing an SLT by age 5 by EPVT score

Note: The red threshold represents 1 standard deviation below the mean, the most commonly applied threshold to the EPVT/PPVT. The two green thresholds represent 1.5 and 2 standard deviations below the mean.

[4] For an example of what an "ideal" RDD would look like from fictional data, see Appendix 3.

Figure 5.2 plots the probability (range 0-1) of children having seen a Speech and Language therapist before sitting the test for each score on the EPVT distribution. There is no visible difference in the number of children receiving support for a possible language disorder observable at the threshold. This is equally true at age 10 (Appendix Figure A5.4). The test was created specifically for this cohort and so had not been standardised or tested for diagnostic potential. This means that no referrals were made based on scores on this test and so no self-selection into scores below the threshold occur motivated by service eligibility.

Due to the normal distribution of the EPVT scores, the central scores have larger sample sizes per group, meaning that outer values can be more skewed by the presence of individual anomalies. The linear regression accounts for the impacts that smaller sample size has on observable discontinuities in outcomes. The impact of discontinuities in SLT attendance at some values in the distribution is tested through restricted sampling, as described in the Stage 2 methodology.

This means that if we observe any discontinuities in outcomes, they are explicitly due to functional differences (or spurious anomalies) and not due to any effects of intervention caused by their score on the EPVT leading to referral.

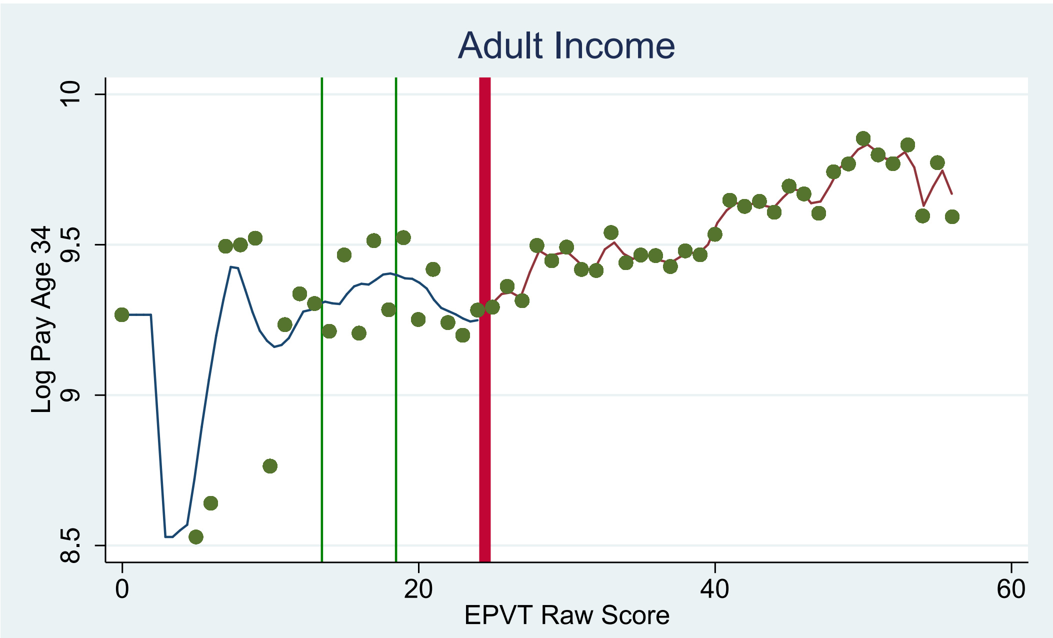


Figure 5-3 Local polynomial of average earnings at age 34 by EPVT score

Note: Threshold values at 1 s.d. (red), 1.5 s.d. and 2 s.d. (green) plotted.

The same plots can be used for later life outcome measures, such as the log of income at age 34. The discontinuities seen here are at much lower scores than any threshold and may reflect small sample size and the influence of large anomalies (those with no/minimal income or a lack of large incomes). Lower sample size, whilst not depicted here, is reflected in the significance of results in any linear regression.

Once again, the regularly used thresholds for identifying impairment don't seem to hold any extra predictive power over the outcome not picked up by the general trend line (improved vocabulary at age 5 correlates to higher incomes at age 34, as expected).

5.7 Stage 1 Results

The results of the Stage 1 regressions are presented in Figure 5.4. All outcomes are standardised to allow comparable coefficient plots.

No outcome has a functional negative discontinuity at the threshold drawn one standard deviation below the mean at the 10%, 5% or 1% significance level. The threshold does not significantly predict lower functional language ability having any impact on any of the reported domains.

There is one positive (theoretically equivalent to a non-result) estimate at the 5% significance level, that children with the identified language disorder earn less than children without the identified language disorder in the income measure calculated over all four timepoints. This result would not be significant if a Bonferroni correction were applied to account for the 15 regressions ran in the Stage 1 analysis.

Across the 15 outcomes, there are a range of positive and negative estimates. There is no category of outcomes which show consistently negative associations, even non-significant ones, with the threshold identification. This suggests that, after accounting for the linear trend, there is no consistent impact of the threshold in either direction, i.e. those below or above the threshold continue to follow the linear trend with nothing changing at the threshold.

5.7.1 Stage 1 Research Question and Hypotheses

1. Can the most commonly applied threshold for the English Picture Vocabulary Test accurately identify functional difficulties requiring intervention?

   Null Hypothesis H_0: $\tau_{24,\,1\text{-}15} \geq 0$

   Alternative Hypothesis H_A: $\tau_{24,\,1\text{-}15} < 0$

At the 10% significance level, we cannot reject the null hypothesis that a one standard deviation threshold does not identify functional language difficulties that would require intervention.

Children narrowly below the threshold commonly used to identify language impairment are indistinguishable from those narrowly above the threshold in later life. In conclusion, if a language disorder is identifiable at the margin by a strictly applied threshold on the English Picture Vocabulary Test, that threshold is not one standard deviation below the mean.

This threshold would not accurately diagnose the presence of a language disorder with any additional accuracy to the general incremental improvements anywhere else along the distribution of scores.

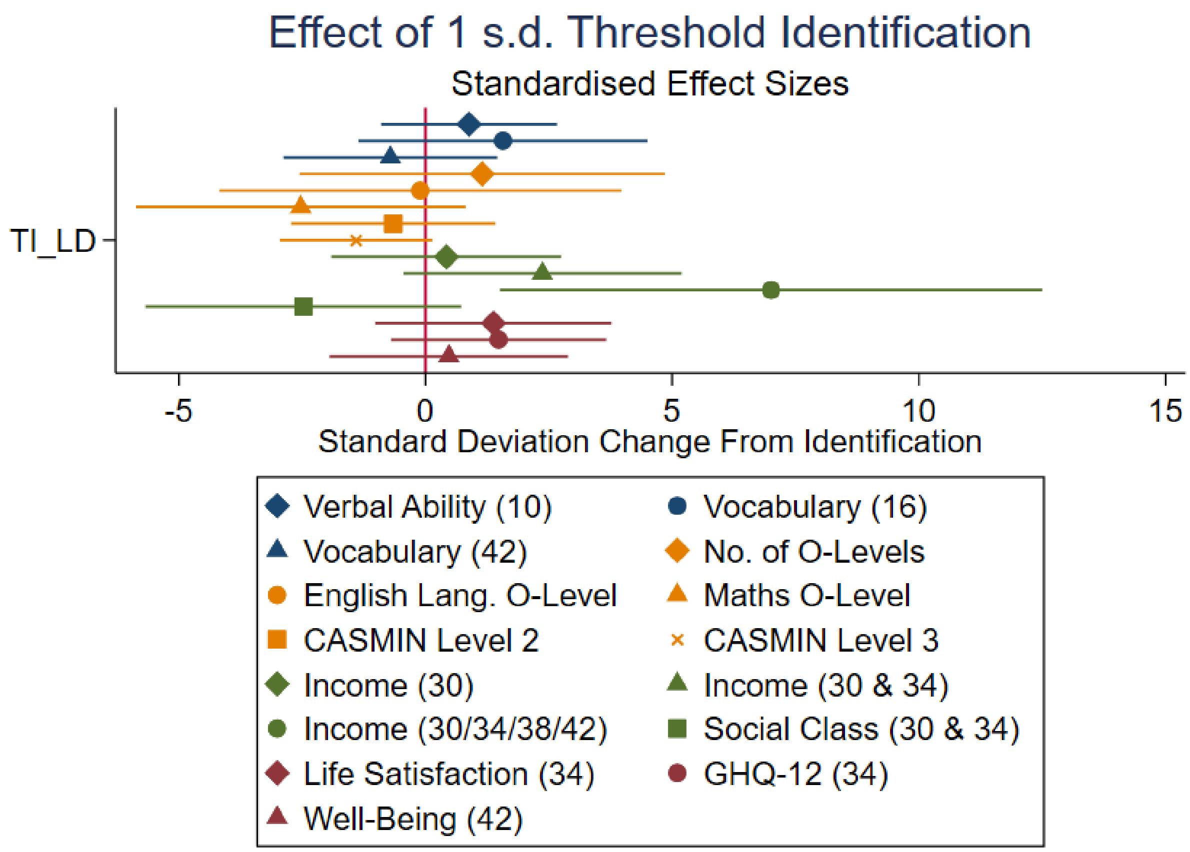


Figure 5-4 Effect of 1 s.d. threshold on each outcome

Note: All estimates of Tau (τ), the marginal effect of using the threshold to identify a language disorder at the threshold, for each outcome. Outcomes are categorised as in Table 5.3, where blue = language, yellow = education, green = occupation and red = health/quality of life. Horizontal lines represent a 95% confidence interval around the point estimate. A 95% confidence interval lying to the left of the 0 line would indicate a successful identification procedure at the margin.

5.7.1.1 Robustness Checks

The results presented in Figure 5.4 are robust to the use of four sub-samples:

- Those who attended a speech and language therapy service before the age of 10
- Those who did not attend a speech and language therapy service before the age of 10
- Males
- Females.

Within each of these four sub-samples, no outcome presented an additional negative impact for those scoring below the identification threshold beyond the general linear trend. Equivalent data to Figure 5.4 is reported for each of the subsamples in Appendices A5.5 to A5.8.

Furthermore, the results presented in Figure 5.4 are not sensitive to changes in the bandwidth for inclusion around the threshold. The primary analysis includes individuals with scores from 20 to 29 (five scores either side of the threshold). Appendices Figures A5.9 to A5.12 reports the equivalent data to Figure 5.4 for an inclusion bandwidth of three, four, six and seven either side of the threshold. All conclusions and hypothesis tests are robust to these changes. Testing of this sensitivity, whilst uncommon, is recommended given the potential for researcher bias in selection (Lee and Lemieux, 2010).

There are clear implications to these results. Clinically, if intervention were allocated to those scoring below the threshold only, and those scoring above the threshold were excluded, this would separate children who have identical expected outcomes. This identification procedure may therefore create an inequality between children at a point where none previously existed. For researchers, this suggests that in previous studies, children with equivalent expected outcomes would be identified as belonging to different language groups. This limits the potential to identify the impact of genuine group differences.

This then raises a simple question for analysis in Stage 2 – Is there any justification for a strict cut-off being applied to language assessments?

If a threshold drawn at a different point could accurately identify difficulties, then that threshold should be the one applied. As discussed in the literature review, there is no consensus on the optimal threshold to be applied that is consistent across studies and language assessments.

Stage 2 addresses this question, utilising all feasible thresholds from 9 to 35 on the English Picture Vocabulary Test.

5.8 Stage 2 Results

In Stage 2, all feasible thresholds between 9 and 35 for their ability to predict functional language difficulties impacting the same 15 outcomes as in Stage 1. This gives a total of 405 estimates (27 thresholds * 15 outcomes) of τ.

From 405 τ estimates, 10.2 estimates (2.5% on the left-hand tail at the 5% significance level) would be expected to be significant purely through Type I error.

A dataset of results from the 405 regressions is created, recording all estimates of τ (as well as our constant, gradient and change in gradient estimates), all standard errors, p values, associated outcome labels and the threshold used.

5.8.1 Discontinuities Estimated by Threshold and Outcome

11 estimates are found to be significant at the 5% level in the negative direction. This is in line with our Type I error expectations.

17 of the 27 thresholds tested predict none of the 15 outcomes significantly, including the 1 standard deviation threshold at 24 tested in the Stage 1 analysis.

A further 9 thresholds predict one of the 15 outcomes significantly.

A threshold identifying children scoring 11 and below as having a language disorder significantly predicted lower outcomes for vocabulary scores at both 16 and 42. This is the only threshold which identified two discontinuities which could indicate the presence of a disorder.

Figures 5.5 to 5.8 summarise the dataset of estimated results. As is visible in these Figures, most estimated Tau parameters (95%) are not significant and there is very little pattern to the results which are significant.

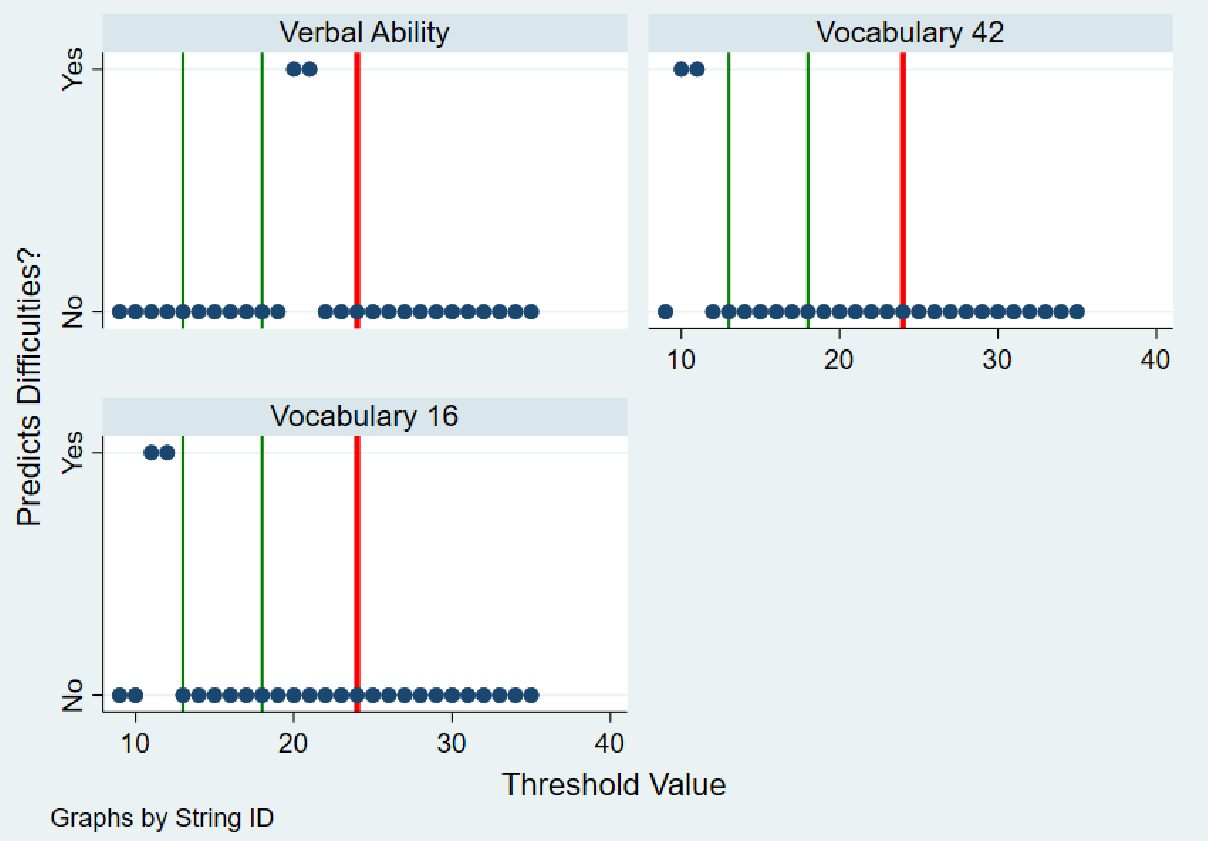


Figure 5-5 Marginal identification effect of each threshold on language outcomes

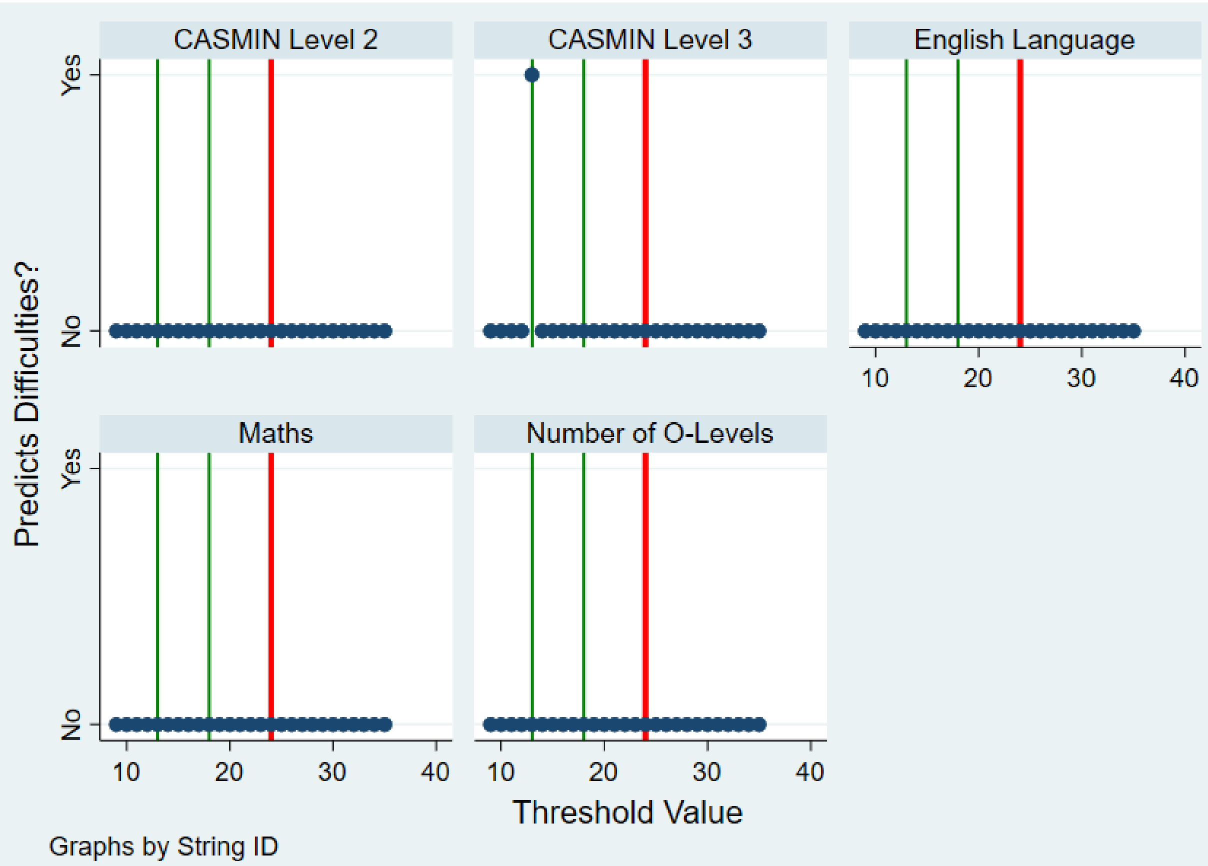


Figure 5-6 Marginal identification effect of each threshold on education outcomes

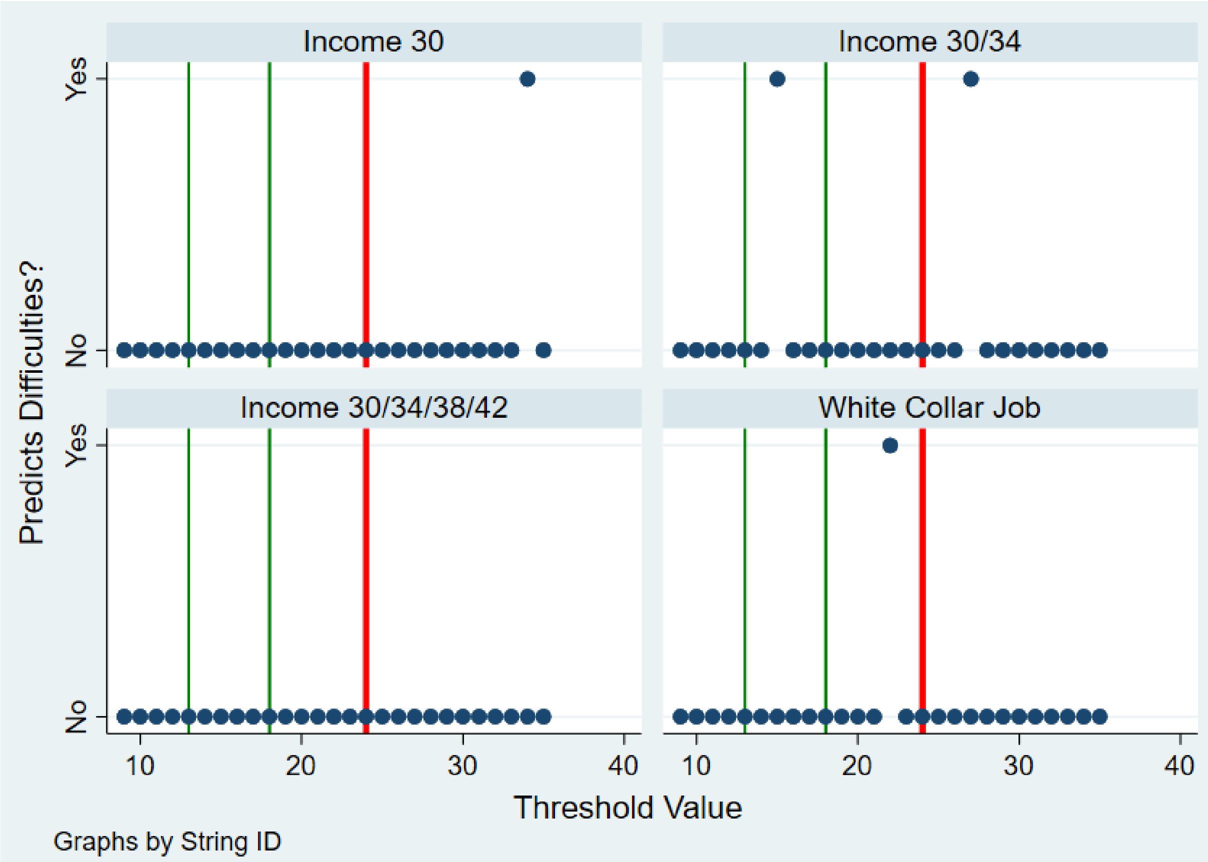


Figure 5-7 Marginal identification effect of each threshold on occupational outcomes

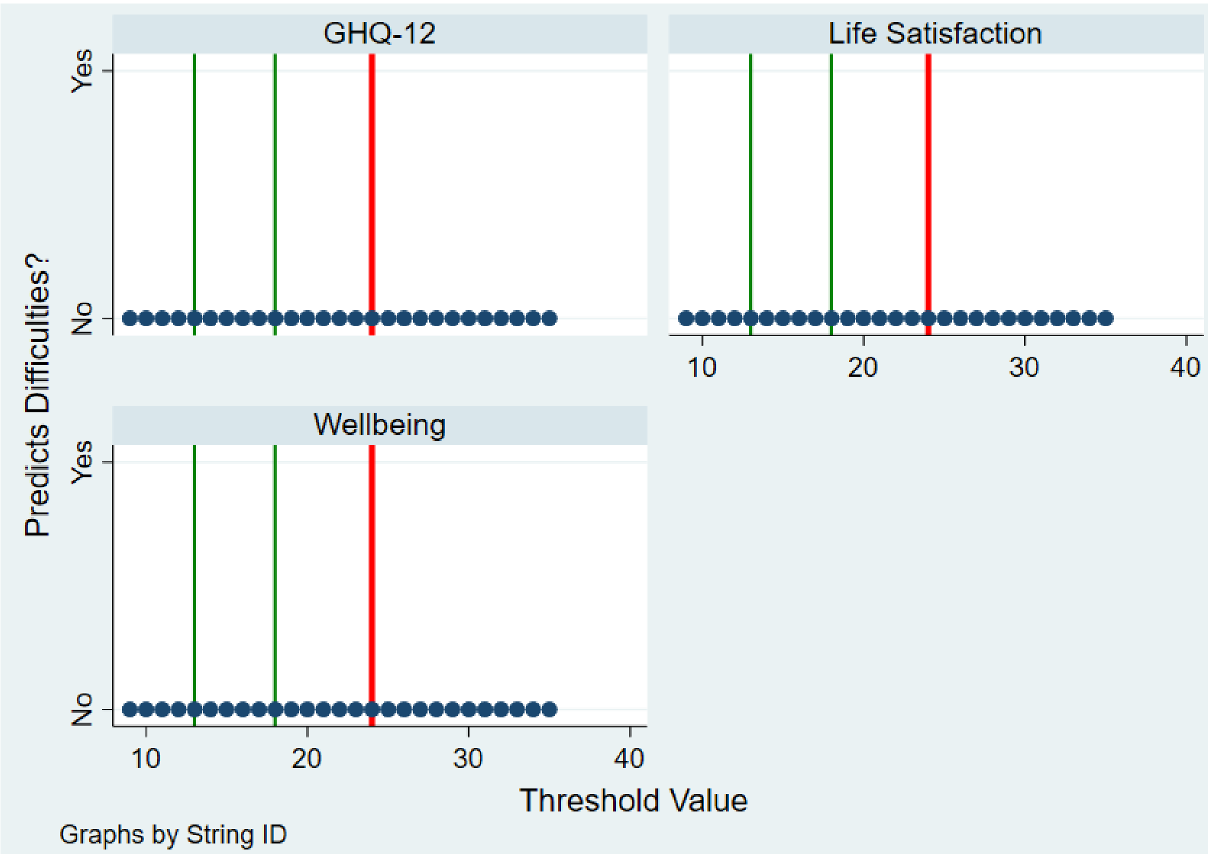


Figure 5-8 Marginal identification effect of each threshold on health/quality of life outcomes

5.8.1.1 p Values

A low p value implies a greater confidence that there is a difference between those identified with the disorder by the threshold (TI-LD) and those above the threshold (TI-NLD). A p value of below 0.05 would denote a significant identification on a given outcome. When averaging the p value of all outcomes for each threshold (see Figure 5.9), no threshold is close to being consistently predictive of all outcomes.

There is no evidence that any threshold can accurately identify a language disorder that impacts a range of outcomes at the margin.

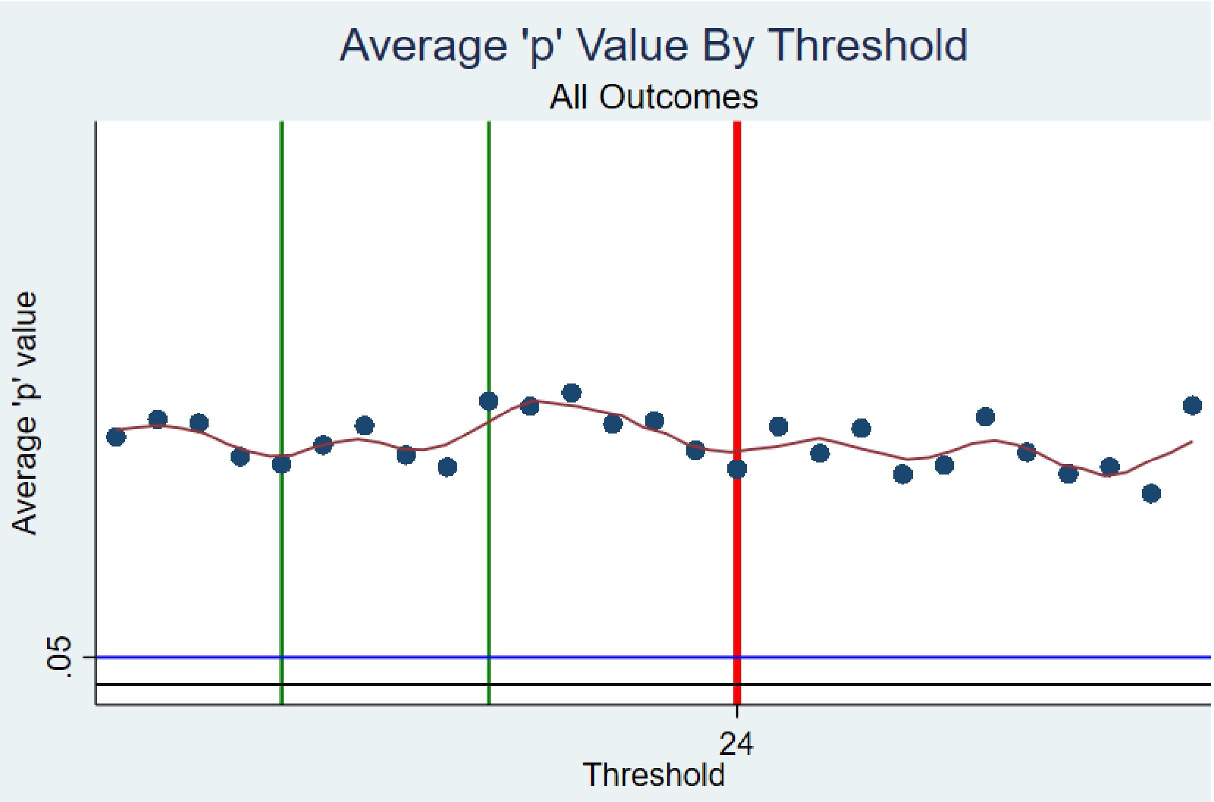


Figure 5-9 p values plotted by threshold choice

Note: The average p value of the Tau indicator of a Threshold Identified Language Disorder, averaged across all 15 outcomes, for each threshold. A local polynomial is fitted to the mean p values. The commonly applied thresholds are highlighted at 1, 1.5 and 2 standard deviations away from the mean (the furthest right-hand value). A data point below the 0.05 line would indicate that a threshold was generally predictive of all outcomes.

5.8.1.2 The Commonly Applied Thresholds

None of the 11 estimates that are significant at the 5% level occur at the commonly used thresholds of 1, 1.5 or 2 standard deviations below the mean.

Furthermore, the average p value of estimates at the commonly applied thresholds are no different to those found at arbitrary thresholds drawn from across the spectrum. This is visible in Figure 5.9 and statistically demonstrable through t-tests. The commonly applied thresholds also perform equally to arbitrary alternative thresholds on Kruskal-Wallis comparisons of medians.

These findings are robust for each individual outcome as well as for each category of outcomes.

In summary, these results offer no justification for the tendency to select thresholds at 1, 1.5 or 2 standard deviations below the mean in favour of any other threshold.

5.8.2 Stage 2 Research Question and Hypothesis

2. Can any alternative threshold applicable to the English Picture Vocabulary Test accurately identify functional difficulties requiring intervention?

   Null Hypothesis H_0: $\tau_{c,\,1\text{-}15} = 0$

No value of $\bar{c}$, the threshold, has predictive power over outcomes 1-15.

   Alternative Hypothesis H_A: $\tau_{c,\,1\text{-}15} \neq 0$

At least one threshold value, $\bar{c}$, is an accurate predictor of at least one of the 15 outcomes.

For this research question, we are testing a universal null hypothesis. In this instance, Armstrong (2014) recommends that a Bonferroni correct be applied to all p-values. Therefore, we are testing whether any discontinuity identified remains significant after applying a Bonferroni correction to account for the number of regressions ran.

At the 10% significance level, we cannot reject the Null Hypothesis that no threshold can accurately identify functional difficulties requiring intervention.

5.9 Results Summary

Table 5-5 Conclusions drawn according to results combined with Table 5.4

		Stage 1 Research Question	
		1 s.d. is effective	**1 s.d. is ineffective**
Stage 2 Research Question	**At least one other threshold is effective**	Strictly applied thresholds can identify language disorders, but a choice is required.	Strictly applied thresholds can identify language disorders. The wrong threshold was previously used.
	All other thresholds are ineffective	A one standard deviation threshold uniquely identifies language disorders.	No strictly applied threshold identifies language disorders.

Table 5.5 updates our hypotheses drawn in Table 5.4. We were unable to reject the Null Hypothesis associated with either of our research questions.

This leads to the conclusion that no strictly applied thresholds could accurately identify the presence of a language disorder at the margin on the English Picture Vocabulary Test within the British Cohort Study.

5.9.1.1 Post-Hoc findings

A threshold drawn at a score of 11 on the EPVT predicted worse vocabulary scores at age 16 and 42 for those narrowly below the threshold relative to those narrowly above. Figure 5.5 shows that two more discontinuities on these outcomes are found at thresholds of 12 and 10 respectively. The universal null hypothesis tested in Section 5.9.2 concluded that a strict threshold at any of these points would not accurately identify a language disorder.

Severely low vocabulary, between 2.17-2.36 s.d. below the mean (scores of 10-12), predicts discontinuously low vocabulary scores in adulthood. This may imply that the lowest 2% of performers on the EPVT at five have significantly poorer later acquisition of vocabulary beyond that captured by the linear trend.

Two caveats are noted to this finding:

Firstly, this is not a strict threshold effect as thresholds at 10, 11 and 12 demonstrate similar effects. Therefore, this finding would not support the application of a strict threshold but may suggest the need for enhanced support targeting language acquisition at the most extreme language disorder severity.

Secondly, this finding must be interpreted with caution as these results are not significant after the application of a Bonferroni correction.

Further research would be necessary to establish whether a particular range of severely low early receptive vocabulary could identify a group of children who are at risk of poor vocabulary development into adulthood. If this finding was supported with alternative assessments used directly in language disorder diagnostics, it could motivate the provision of enhanced interventions for this group which specifically target language (e.g. Boyle et al., 2007), as opposed to broader educational goals.

5.10 Discussion

5.10.1 Benefits and Drawbacks of Language Assessments and Thresholds

There are clear benefits to the use of language assessments. They capture elements of a child's underlying language ability and can help to identify the children at the lower end of the distribution who are most likely to require support. The provision of percentiles and norm-referencing is a useful reference point for speech and language therapists who typically see children who perform below their age expectations.

The value of specific thresholds applied to scores on these assessments is more contested. The current chapter seeks to analyse the value of thresholds applied to these tests, commonly used to indicate the presence or severity of a language disorder (Wiig et al., 1992). Specifically, the analysis compares children who score close to the threshold value, testing the diagnostic capability of thresholds at the margin.

The results of these analyses clearly demonstrate that no strictly applied threshold on the English Picture Vocabulary Test could identify the presence of an impactful impairment. Children narrowly above any threshold (TI-NLD) perform similarly to those narrowly below that threshold (TI-LD), regardless of the choice of threshold.

Furthermore, the thresholds which are most applied within the literature (including those specifically applied to this assessment within this dataset) have no greater marginal predictivity of later difficulties than other thresholds drawn arbitrarily across the distribution. Wiig et al.'s (1992) summary of these thresholds indicating mild, moderate and severe impacts of a language disorder are not evidenced at the margin.

A strictly applied threshold within this dataset, wherever it was drawn, would separate children who have equivalent life chances. In practice, this may lead to the provision of support to one group over another where no justification exists.

Applying a threshold strictly has some benefits. Knowing that a certain proportion of the population will qualify for a policy enables resource management. RCSLT (2018) estimate that around 17,000 speech and language therapists work in the UK, two thirds of which work with children. Application of a strict threshold nationally or regionally would allow for the estimation of caseloads and necessary recruitment. Resources are typically stretched given the prevalence of Speech, Language and Communication Needs (SLCN) and simple diagnostic processes may save some time.

When thresholds are assumed to be broadly effective, this process is convenient and justifiable. However, there is an ethical issue arising from the results presented in this chapter. Children are allocated access to services according to need. The range of poor expected outcomes referenced in most of the literature signify the needs of this population. The results of this chapter suggest that the level of need cannot be differentiated accurately by a strictly applied threshold on the EPVT.

Just 15% of the people surveyed in the Bercow Ten Years On (Bercow, 2018) report believed that the right level of speech and language therapy was available in their area. This suggests that the threshold selected is unlikely to be applying the differential treatment of equal children at a level indicating typical development. Thresholds drawn below the mean are likely to exclude some children who have a genuine language disorder and this risk increases as the threshold becomes more severe (Spaulding et al., 2006).

5.10.2 Strengths and Limitations of the Data

The EPVT set of plates (words tested) within the British Cohort Study are unique to this dataset, combining the pre-school and school years plates. They were not designed for a specific diagnosis. This test, using a within-sample standardisation, is unlikely to be diagnostically accurate, relative to a language assessment specifically intended for diagnostics. However, the testing of every feasible threshold to maximise the diagnostic potential of the measure used is a benefit in this regard which is not present in other assessments. Whether the results presented can be replicated with a test and threshold intended for diagnosis requires further research.

Whilst a test with high sensitivity and specificity (high rate of true negative and true positive) may perform more accurately at the margin, there was not a dataset available to test this in the same way. An important unique feature of the data was the lack of referral from the EPVT scores. This allowed the estimation of what differences would naturally develop between children either side of a threshold in the absence of intervention. This natural experiment informs the conclusion that the test would not have been diagnostically useful at the margin. Within cohorts that use diagnostic tools, referrals that are made from the results would bias the assumption of equality either side of the threshold. The estimates produced would contain both the threshold identification effect as well as a potential intervention effect, complicating the evaluation. Furthermore, the reporting of results within a norm-referenced framework to parents may influence future parental investments in the child, further biasing the estimated impact of the threshold itself for identifying between-child differences.

5.10.3 Strengths and Limitations of the Method

Regression Discontinuity Designs are highly useful for estimating the impacts of policy-relevant thresholds. This design had not been previously applied to the field of speech and language sciences, to the author's knowledge.

The application presented is unusual as it estimates the difference in outcomes in the absence of a policy, as opposed to estimating the effect of a policy. Usually, a threshold will only have impact through differential treatment. In this application, the "treatment effect" is the hypothetical presence of a language disorder, as identified by the threshold.

The looped design is a further innovative use of the RDD method, not previously applied to any field to the author's knowledge. In this instance, it allowed the testing of all feasible thresholds and their direct comparison. For further applications, this method is an alternative to the Imbens (2004) approach of testing a number of random "placebo" discontinuities to demonstrates the unique effect of a policy-relevant threshold.

Alternative methods could test the same hypothesis. For example, a t-test could be used to compare groups either side of a given threshold and looped to test all feasible thresholds. The current technique is preferred as it allows the estimation of a linear trend, separating the impact of moving across the threshold from the impact of a unit increase in the language score. Furthermore, the Lee and Lemieux (2010) guide to applying an RDD provides guidance on factors such as bandwidth for inclusion and robustness checks.

This method would be suitable for further assessment of the research questions. It would be worthwhile to test whether the poor identification at the margin within the British Cohort Study extends to other outcome measures and other language assessments. More broadly, a Regression Discontinuity Design could be applied to similar fields where health conditions operate in terms of both diagnosable traits and more continuous measures of severity.

5.10.4 Aims of the thesis

The following broad aims were outlined within this thesis.

1) To increase understanding of the length and breadth of long-term impacts of early speech and language development.

Language, Educational, Occupational and Quality of Life or Health outcomes are used as the indicators of functional difficulties, measured between the ages of 10 to 42, arising from poor initial receptive vocabulary, as measured at age five.

2) To utilise historical data alongside econometric methodologies previously unused in the field.

Regression Discontinuity Designs had not been previously implemented in speech and language sciences, or using the British Cohort Study 1970, to the author's knowledge. However, RDDs are a common technique in a range of fields including Education, and this technique in increasing in popularity due to its simplicity and intuitive relevance to many policy questions (Villamizar-Villegas et al., 2020).

3) To make appropriate recommendations for research and clinical practice.

The results presented suggest that no threshold would have accurately identified the presence of functional language difficulties at the margin. These results are only generalisable to the specific test analysed, the bespoke English Picture Vocabulary Test within the British Cohort Study. If these results were replicable in assessments used diagnostically, this suggests that applying strict thresholds in allocating services would induce inequalities between children with equal levels of need at the margin.

Furthermore, the results suggest that researchers should be wary of identifying groups in longitudinal studies as having language disorders on the basis of strictly applied thresholds as they are likely to contain children in the identified and unidentified groups who are equal. Reporting of the sensitivity of results to changes in the threshold applied is recommended, as lower thresholds may increase the observed group differences and threshold choice remains subjective. Any of the common thresholds (1, 1.5 and 2 standard deviations below the mean) have precedence in the literature and the choice between these is likely to impact the findings.

The presence of differing levels of severity of a language disorder is widely recognised (Norbury et al., 2016; Bishop et al., 2017; Wiig et al., 1992; Law et al., 2017a). The analysis presented suggests that language difficulties are linearly graded along the spectrum, although a specific point at which the severity increased notably was not identifiable within this language assessment. Whilst these results are limited to the measures available, they represent evidence that language ability and language difficulties are best measured continuously, implying that service provision should be graded as much as possible according to a continuous spectrum of need. This means excluding as few children as possible from support given the high cost of unresolved communication difficulties (see Chapter 2).

Practically, it is very difficult to provide a spectrum of support across the whole distribution. The focus should be to minimise the exclusion of children at the lower end of the distribution, as lower language scores *tend* to lead to poorer outcomes. In screening terms, this means using the mildest feasible thresholds across multiple tests. Whilst this will still mean a sub-optimal inclusion and exclusion of equivalent children arbitrarily, the intention should be to minimise the number of children with severe language disorders not diagnosed due to a stochastic random error. Fundamentally, this means more children would be in receipt of prolonged observation, if not intervention. This suggests either greater resources are required, or the spreading of available resources through options such as group-based and school-based interventions which do not require formal diagnoses and strict thresholds.

Significant differences in this level of need were not identifiable through a strictly applied threshold and therefore strictly applied thresholds may represent a poor criterion for service eligibility.

6 Chapter 6: Discussion

6.1 Summary and Thesis Aims

This thesis explored the life experiences of children who experience speech, language and communication needs in the first seven years of life. Two datasets were examined to thoroughly investigate the outcomes for these children, relative to typically developing peers, across multiple spheres of life. A complex picture is painted in two ways.

Firstly, the outcomes for children who do not develop strong speech or language skills are not universally impacted. Whilst educational and occupational outcomes are typically lower within affected groups, the results for health and non-cognitive outcomes are more mixed although no results report significantly greater outcomes for the affected group.

Secondly, even where significant associations are found, the maximum and minimum achievements are the same for all speech and language groups. Children with low vocabulary or early speech difficulties can go on to achieve post-graduate qualifications and can receive earnings within the highest income categories. The analytic approaches taken model conditional expectations, and so the negative associations noted reflect lower life chances, not determined outcomes.

The following section summarises the research questions approached in this thesis and discusses the implications of the results.

6.2 Overview of Analyses, Research Questions and Contributions

6.2.1 Chapter 1

The 'Background' chapter outlined the broad objectives of the thesis and proceeded to summarise the essential literature relating to childhood speech and language development. The complexity and unpredictability in the formation of the important skill of communication was a consistent feature of the literature reviewed. Many speech, language and communication needs present in the pre-school years resolve spontaneously, although the potential for spontaneous resolution decreases with age (McKean et al., 2017a). Whilst many risk factors are associated with the probability of either speech or language needs developing, predictive risk models remain weak in their capacity to predict which children will not develop typically (Camacho et al., 2019; Reilly et al., 2010).

Relatively few studies explore the broader achievements of children with SLCN beyond their school years. A number of key papers are identified in order to develop hypotheses for the analyses conducted in later chapters. These are referred to directly in response to specific research questions.

6.2.2 Chapter 2

1) Are early speech and language development profiles observable in the Newcastle Thousand Families Study?
   a) How does the observed prevalence of childhood Speech, Language and Communication Needs (SLCN) compare to the current literature?

This chapter identifies 18.2% of children with some form of SLCN between the ages of two and seven. Whilst this is higher than the prevalence rates reported in cross-sectional studies (Law et al., 2017a; Norbury et al., 2016; Tomblin et al., 1997), this is likely to be due to the length of time allowed for an identification. The regular observations for five and a half years allow for the identification of more transient difficulties, especially given the unstable trajectories of language within early years (McKean et al., 2017a).

Approximately half of the SLCN cases were resolved by school entry, whilst half persist to at least this age (a very small minority emerge after school entry). Literature suggests that around 70% of 'late talkers' resolve their delays by age four (Reilly et al., 2010). Of the 70%

who resolve, Rescorla (2011) find that they still perform worse on language assessments on average than the group who never experienced a delay. The subjective judgement of the health visitors within the Newcastle Thousand Families Study may therefore reflect the lower likelihood of resolution being specifically noted.

The two most relevant risk factors to the current study, gender and socio-economic status (measured here by the father's occupational class), are predictive of SLCN status, as is consistent with the literature (Tomblin et al., 1997; Hart and Risley, 1995). The two cited papers from the 1990s have been replicated in modern studies (Leonard, 2014; Lohndorf et al., 2018) which demonstrates the enduring nature of these associations. This suggests that a major corrective effort would be required to address these inequalities.

2a) What are the effects of SLCN on academic trajectories from the 11+ secondary school examinations through to the maximum academic attainment observed by adulthood?

Whilst many studies identify an attainment gap at school entry and within primary school (Young et al., 2002), persistence of this gap until completion of education is often assumed rather than formally tested. Durkin et al. (2009) demonstrate gaps in achievement for individuals with language impairment at GCSE age and Conti-Ramsden et al. (2018) have since demonstrated the existence of achievement gaps in further education settings.

The analysis in Chapter 2 finds that people with a history of SLCN have poorer average outcomes in educational settings. They have lower 11+ exam scores, a lower probability of obtaining any form of academic qualification, and a lower probability of obtaining each subsequent higher level of academic qualification. This is supported by the expectations drawn from the literature before the analysis was conducted and is consistent with the findings of Conti-Ramsden et al. (2018) regarding further education, a study published after the present analysis was initially conducted.

2b) What are the effects of SLCN on earnings in adulthood?

SLCN predicts lower household income at age 50, although not at age 62. The prior result is consistent with the hypotheses drawn from the literature, suggesting that communication skills are valued in the labour market (Archer and Davison, 2008; Lowden et al., 2011), that SLCN predicted difficulties with finding employment (Law et al., 2009; Whitehouse et al., 2009) and those in employment were in lower skilled occupations (Carroll and Dockrell, 2012; Conti-

Ramsden and Durkin, 2012; Johnson et al., 2010). However, the latter finding was contradictory. Conti-Ramsden et al. (2018) and Records et al. (1992) both find similar levels of income between identified SLCN and typical groups, although these studies are both conducted in early adulthood.

There was a high retirement rate (67%) amongst respondents at age 62, reducing the variation in incomes. This limits the interpretation of the lack of association at this age, as the research question was intended to measure earnings from employment. Conclusions relating to earnings were drawn more thoroughly in the following chapter, as this question was crucial in understanding the economic and personal implications of early speech and language development.

2c) What are the effects of SLCN on health in adulthood?

SLCN was mostly unrelated to health outcomes in adulthood. One finding associated SLCN with psychiatric risk at age 50, although this finding was driven by those with resolving SLCN. There was no theoretical reason for the resolving group to be impacted rather than the persistent group. Therefore, the overall conclusion drawn was that SLCN did not significantly predict health outcomes within this sample. The use of an objective health measure, grip strength at 62, supports the validity of this conclusion. Furthermore, there were no associations found with the CASP-19 measure of quality of life for individuals approaching retirement. These results contradict the small body of literature reviewed in Chapter 1 regarding mental health and health-related quality of life (Law et al., 2009; McKean et al., 2017b; Clegg et al., 1999; Le et al., 2020a; Le et al., 2020b).

2d) What are the effects of SLCN on the OCEAN non-cognitive skills as measured in adulthood?

Lower rates of Openness to Experience and Extraversion are found within the SLCN sample, although statistically equivalent rates of Conscientiousness, Emotional Stability and Agreeableness are also found. There was relatively limited literature from which to draw hypotheses (literature synthesised in Table 1.2) and these results of Chapter 2 represent the only directly measured associations in the literature.

The lower rates of extraversion are consistent with intuitive and literature-drawn expectations (Lindsay and Dockrell, 2000; Durkin et al., 2017; Wadman et al., 2008).

The lower rates of openness to experience were pre-empted through the lower rates of further education attendance (Conti-Ramsden et al., 2018) although no prior research existed.

Conscientiousness was largely considered only in terms of hyperactivity in young children (Levickis et al., 2018b; Westrupp et al., 2020) as an indicator of low attention span. The lack of significant group differences imply that this deficit is not persistent throughout life, and that people with SLCN showed no deficits in work ethic or attention to detail.

Facets of agreeableness had been widely researched in the past, with negative associations noted for aggression in early years, particularly for males with SLCN (Brownlie et al., 2004; Yew and O'Kearney, 2015) as well as problems with co-operative peer relationships (Durkin and Conti-Ramsden, 2010; Rose et al., 2018). Positive associations had been noted for prosociality (Toseeb et al., 2017; Wadman et al., 2008; Durkin and Conti-Ramsden, 2010; Toseeb et al., 2020; Toseeb and St Clair, 2020) which may explain the lack of overall differences observed in agreeableness within this sample.

No significant differences were observed in emotional stability, contrary to the expectations drawn from the literature. Anxiety, self-esteem and shyness are all elements of emotional stability anticipated to be poorer within the SLCN group (Beitchman et al., 2001; Fujiki et al., 2002; Botting et al., 2016; Durkin et al., 2017).

The core contributions of Chapter 2 are the generation of an identifier of SLCN, including measures of persistence and severity, and the subsequent association of SLCN with multiple levels of educational attainment, and with non-cognitive skills in later adulthood.

6.2.3 Chapter 3

1) Does a child's vocabulary predict their earnings in adulthood?

We find that receptive vocabulary at age five significantly predicted higher earnings from age 30 to 42. A standard deviation increase in vocabulary correlated to around a 3% increase in annual earnings. This supports the majority of the expectations drawn from the literature, although contradicts findings from Conti-Ramsden et al. (2018). We conclude that these income differences are likely to only be observed later in adulthood and that the Conti-Ramsden et al. study did not extend far enough into an individual's career to establish their maximum earnings, as many individuals would only recently have left education. This hypothesis is supported by the general literature regarding income differentials (Grawe, 2006) and suggests that future research should use datasets which extend to at least age 30 (Gregg et al., 2014).

2) Does performance in the education system mediate any relationship between early vocabulary and earnings?

We do not find evidence of a mediation effect. This result was unexpected based on the prior literature as both paths within the mediation, vocabulary to education (Durkin et al., 2009; Young et al., 2002) and education to earnings (Dolton and Vignoles, 2000), have been previously well established.

3) Are non-cognitive skills developed in later childhood mediators of any relationship between early vocabulary and earnings?

We do not find evidence of a mediation effect. Locus of control at age 10 was the only non-cognitive measure tested within this analysis, which may explain the lack of an effect as there is only weak evidence of an association between either language and locus of control (Markham et al., 2009) or locus of control and earnings (Almlund et al., 2011) in the prior literature, as not many studies previously test these relationships.

The core contribution of Chapter 3 is the evidence of lower earnings amongst individuals with lower receptive vocabulary in childhood. The size of this correlation is discussed, as is the persistence of this association between ages 30 to 42.

6.2.4 Chapter 4

Chapter 4 employs a Oaxaca (1973) form of the Oaxaca-Blinder decomposition, segmenting the earnings gap between two groups (low vocabulary compared to middle and high vocabulary groups separately) into an endowment effect, relating to mean levels of covariates, a coefficient effect, relating to differences in the value of these covariates in the labour market between groups, and an interaction effect to account for the prior two effects occurring simultaneously.

1) Are there significant differences in the covariates of low and high vocabulary groups?

We find that covariates drawn from socio-economic status, cognitive skills, non-cognitive skills or educational attainment are all significantly different in mean levels between vocabulary groups.

2) Do these differences lead to an earnings gap between low and high vocabulary groups?

A significant 'endowment effect' is found at the 1% significance level. In particular, the covariate denoting educational attainment is attributed a large degree of explanatory power (around a third of the observed earnings gap) within this model. Cognitive skills, measured as motor-visual co-ordination, maths ability and reading ability, are also significantly predictive of earnings differentials between the low and high vocabulary groups.

The finding that education is crucial in explaining the earnings gap between vocabulary groups does not necessarily contradict the findings of Chapter 3. Firstly, the earnings gap here is not along a continuous distribution, but an average within distinct groups, reducing the measurement error. Secondly, the decomposition method has to attribute the difference to one of the available variables, rather than to an error term. Thirdly, the mediation effect could only be a proportion of the overall effect, including a direct pathway which is not present in Chapter 4.

3) Does the vocabulary ability of the individual moderate the labour market value of these covariates?

No significant coefficient effect is observed, indicating that vocabulary does not alter the returns to the set of skills, characteristics and qualifications analysed.

6.2.5 Chapter 5

Chapter 5 extended the counter-factual approach of the thesis by removing the impact of covariates altogether. The previous chapters had shown that measures of early communication can predict later outcomes, but that differences in later outcomes did not appear to be especially sensitive to any certain points of distribution.

1) Can the most commonly applied threshold for the English Picture Vocabulary Test accurately identify functional difficulties requiring intervention?

A threshold was applied at one standard deviation below the mean in addition to a linear regression. Being below the threshold significantly predicted none of the 15 tested outcomes beyond the general linear trend. This implies that this threshold was not an accurate method for identifying language disorders at the margin.

2) Can any alternative threshold applicable to the English Picture Vocabulary Test accurately identify functional difficulties requiring intervention?

Thresholds were tested at every integer score from 9 (2.5 standard deviations below the mean) to 35 (the mean), covering the full spectrum of feasible thresholds for identifying language disorders. None of the thresholds tested predicted more than two of the 15 outcomes beyond the general linear trend.

In conclusion, no threshold applied to the English Picture Vocabulary Test could accurately differentiate children either side of the threshold. Strict application of thresholds in research or policy environments will separate identical children into differently labelled groups.

Selection of a particular threshold on this test would not be justifiable on grounds of diagnostic accuracy. Thresholds at 1, 1.5 and 2 standard deviations below the mean are no more predictive of outcomes than arbitrarily drawn thresholds across the spectrum.

This type of analysis has not been previously conducted to generate hypotheses. The use of these thresholds in policy documents implies an expected significant result at the commonly used thresholds, although awareness of measurement error (Jongsma, 1982) would greatly reduce this expectation.

The core contributions of Chapter 5 are the statistical evidence that strictly applied thresholds do not effectively identify functional language difficulties which impair life outcomes at any margin on the English Picture Vocabulary Test, and that commonly applied thresholds do not perform any better than arbitrarily drawn thresholds across the entire distribution of scores.

6.3 Strengths, Limitations and Contributions

6.3.1 Aims of the thesis

To reiterate the focus of Chapter 1, three broad aims were set out in the thesis introduction:

4) To increase understanding of the length and breadth of long-term impacts of early speech and language development.
5) To utilise historical data alongside econometric methodologies previously unused in the field.
6) To make appropriate recommendations for research and clinical practice.

The research questions above seek to achieve aim one, via the approach set out in aim two and to make conclusions which are applicable to achieving aim three.

The following section discusses the relative strengths and weaknesses of the thesis in achieving these aims.

6.3.2 Cohort Data

Two sources of data are employed within this thesis. These are both cohort studies followed longitudinally, meaning that early skills can be captured alongside relevant risk factors present in early life. The datasets chosen for this study are selected specifically for their age, allowing the extension of outcomes beyond those previously modelled in the literature.

The age of these cohorts is their greatest strength. The Newcastle Thousand Families Study (1947) is just one year younger than the world's longest running birth cohort (Wadsworth et al., 2006), and the first population study to contain information on speech and language development.

Without evidence of effects into adulthood, as is provided by this thesis and a few key papers reviewed in Chapter 1, it is hard to make an economic argument for providing early support to children with SLCN or lower early communication skills. Evidence of differences in earnings to age 42 and 50, for example, or differences in personality at age 62 serve to highlight the true long-term impact associated with early indicators of development.

In turn, the age of these cohorts is also the most significant weakness of the thesis. Cohorts born decades ago are generations apart from those impacted by current policy and research. Inferences relating to SES, identification procedures, education and labour markets may be out of date. However, this thesis makes a considerable effort to utilise variables which reflect the context of the era, whilst remaining applicable to the modern day. For example, the use of descriptive data to measure speech and language development, followed by the English Picture Vocabulary Test, a standardised measure which remains popular, or the use of CASMIN levels to reflect educational milestones across multiple education systems.

These samples are both prospectively gathered, meaning the data is collected at the time and not asked retrospectively. This has a number of benefits over alternative methods, including the accurate modelling of speech and language development. Without this, it’s possible that resolving difficulties would be forgotten about, or underestimated. In each analysis chapter, observed logged income is estimated in a linear regression. This is preferred to the approach taken in Marsh et al. (2010), where income was predicted using an extrapolation technique on the basis of education outcomes. Within Chapter 3, lower levels of education associated with lower receptive vocabulary do not significantly mediate the differences in earnings (although this is impacted by the approach taken to dealing with missing data within the chapter). This form of analysis is not possible when using an extrapolation of education to produce the predicted income measure, as Marsh et al. (2010) inherently assume a 100% mediation effect in their modelling approach.

A cross-cohort approach allows for greater confidence in the validity of findings, providing evidence that any associations do not occur exclusively in one area or era. The use of a second cohort also allows for areas of relative strength to be maximised in the targeted research questions. For example, the British Cohort Study had much more reliable data on income in adulthood, whereas the Newcastle Thousand Family contains a reliable measure of each of the OCEAN characteristics. Given that approximately 80% of data analysis time is spent cleaning and preparing the data (Dasu and Johnson, 2003), these cohorts were carefully selected to allow for multiple research questions to be analysed within each cohort. As much information is extracted out of each cohort as possible, including the use of multiple outcomes and methods in each.

The two cohorts contain much larger samples than those possible from first-hand data collection or clinical population sampling. Larger samples improve the generalisability of conclusions drawn about the life experiences of those with speech or language disorders.

However, there are significant problems with sample attrition within the British Cohort Study 1970 analyses, as outlined in Chapter 2, meaning that the conclusions drawn are not demonstrably generalisable to the broader population. Children with receptive vocabulary difficulties that persist may be less likely to continue to respond to lengthy questionnaires throughout their adult lives due to lower comprehension ability and literacy. There is a higher average receptive vocabulary ability in the remaining cohort than the initial cohort within Chapters 3 to 5, suggesting that this has a clear impact. The use of multiple waves, assessments and questionnaires in the data, as is necessary for an epidemiological approach, exacerbates this problem and leads to a relatively small sample size compared to other research within this cohort. The Centre for Longitudinal Studies clearly recognise this issue as the more recent cohort, the Millennium Cohort Study, oversamples in populations with higher risk of attrition. This is a drawback of using historical data, that many of the approaches taken to deal with attrition were developed within the past 50 years, and so they were not applied at the initial sampling for the British Cohort Study 1970.

Chapter 2 has less evidence of this difficulty, with the proportion of individuals with SLCN remaining consistent across data waves. This may reflect that only two data collections occurred in adulthood. Anecdotally, this may also reflect the pride that members of the study have expressed in being a member of a local cohort which has produced valuable research. However, the locality of this sample is a separate limitation to the generalisability of results. The labour market, for example, in Newcastle-Upon-Tyne is unlikely to be reflective of any broader population, particularly from a previous era and so conclusions about the value of human capital skills are not generalisable.

The historical context of these cohorts is relevant to all conclusions drawn. Social inequalities in childhood skills change over time as inequalities between parents change (Blanden et al., 2013). The Newcastle Thousand Families Study is likely to report relatively low levels of within-cohort inequality as it only contains individuals from one area.

Maternal nutrition in pregnancy, and early childhood nutrition, were relatively equal in the Newcastle Thousand Families Study due to post-war rationing (Pearce et al., 2005). Contextual factors such as this may impact the expected health inequalities in childhood and later life. This is one small example of how the life experience of individuals born in 1947 or 1970 are vastly different from those of individuals born in 2020. This is a limitation of any study intending to make recommendations for younger generations than those in the sample, although this limitation is exaggerated by the age of this study. Awareness of the importance of speech and language has increased, as has the recognition of the importance of disabilities and educational needs more generally. Therefore, the policies and services available to disadvantaged groups change over time.

The inception of these cohorts spans a 23-year period in the middle of the 20th century, and the most recent data collections used fall a decade into the 21st century. This is an intentional decision, as outlined in the second broad aim of the research. More recent cohorts would be more appropriate for research questions relating only to school outcomes or to modelling the stability of language profiles over time, but the two cohorts selected were the best available for exploring the life experiences of individuals.

6.3.3 Measuring Speech and Language

This thesis captures verbal communication in slightly different ways within each chapter.

One of the fundamental contributions of this thesis is the revival of the speech and language development data recorded in Morley et al. (1955) and Morley (1965), and the use of thousands of original dated records to parameterise SLCN within a cohort from 1947. This data is now available to access for all users of the Newcastle Thousand Families Study.

The variables generated are products of the methodology used in the data collection. The regular observations over five and a half years are a huge strength, allowing for a broad picture of the child's difficulties to be created and a low probability that any significant issues would go undiagnosed. However, there are potential biases present. For example, reference is occasionally made to the regional accent of the child, implying that the health visitors' perspective of what constitutes a speech disorder may differ in some areas from those outlined by Morley et al. (1954). Where issues are explicitly referenced, they can be treated according to the definitions outlined in Chapter 1, but where information is more limited, the health visitors' concept of SLCN which we do not observe may bias their identification in unanticipated ways.

The use of the 'frequency of negative observations' variable was used as a proxy for severity and persistence (how often did the health visitor make a note on the child's difficulties). However, the results presented apply to the SLCN and persistent SLCN identifiers, as this variable is less reliable. For example, some children had less observations in total, and without knowing the cause of this, it is hard to infer that the number of negative observations is not a stronger reflection of attendance than communication.

Since 1949-54, when this data was collected, the field of research into speech development has expanded significantly. A major change is the separation of speech and language into distinct elements of verbal communication (see Section 1.2.2), meaning that the measures of speech and language in the Newcastle Thousand Families Study does not actually separate the two clearly.

The speech and language data within the Newcastle Thousand Families Study is unique within the literature. The data is qualitatively detailed and reflects a historic study within the field of speech and language sciences. However, the uniqueness of the data is also a drawback as it limits the scope for comparison with the broader literature.

These limitations motivated the subsequent analysis chapters, and the use a widely recognised measure of a specific type of language difficulty: The English Picture Vocabulary Test and receptive vocabulary.

Different versions of the Peabody Picture Vocabulary Test (Dunn, 1965) have been revised and translated, and the measure is widely used in research and clinical settings, albeit with inconsistent application of thresholds (Dunn and Dunn, 2009; Law et al., 2009; Roulstone et al., 2015).

The continuous measure, ranging from 0-56 is an important strength of this measure relative to the prior analysis. This allows estimation of the changes associated with improvements in vocabulary, rather than simply a (usually inaccurate) binary indicator of whether a child have a language disorder or not.

The fact that the test only has one scale, measuring receptive vocabulary, and no other scales capturing other facets of language is a limitation. Whilst receptive vocabulary correlates with many other forms of language, including expressive vocabulary, we are unable to make confident inferences about the value of language ability more broadly. Similarly, we are unable to compare the relative importance of different facets of language. This would be a useful reference point for the discussion of findings relating to earnings and non-cognitive skills presented, as there would be different implications for the underlying mechanisms. For example, if earnings were primarily related to expressive vocabulary, this would lead to different recommendations and conclusions. The conclusions drawn are based upon receptive vocabulary, interpreted as an indicator of broader language ability. Although a receptive-only difficulty could impact productivity alone in occupations with literacy demands, the reference made to communication skills by employers is likely to refer primarily to expressive ability.

The measures used to capture speech and language demonstrate significant strengths and limitations. No measures are taken exclusively at an age where spontaneous resolution is likely, allowing for a relatively accurate identification of persistent difficulties. However, neither cohort measured these difficulties in a replicated manner, limiting the potential for like-for-like comparisons with effect sizes reported in the literature.

6.3.4 Measuring outcomes in life

Outcomes are compared in multiple areas of life between children with and without SLCN, and later for children with different levels of receptive vocabulary. These outcome measures extend further into life than in any prior literature. The most proximal outcomes are measured at ages 10 and 11 years old, in Chapters 5 and 2. The most distal outcomes are measured up to age 62, in Chapter 2. This represents over half a century of life experience from which to draw conclusions. Furthermore, outcomes are drawn from education, occupation, language, health, quality of life and non-cognitive skills. This broad approach enables the nuanced conclusions summarised above.

6.3.4.1 Occupational

The measures of earnings within these samples are limited slightly by the categorical measure, as is the case with most survey data. This leads to a lower level of variation than would be present with a continuous variable. This thesis presents the findings associated with the midpoint estimation in the main text, although two alternative procedures produce equivalent findings (interval estimation and Singh-Maddala distribution), as discussed in the relevant chapters. The Conti-Ramsden et al. (2018) paper is also based upon a categorical income measure, and so the novel findings of lower earnings within SLCN groups are not driven by this limitation.

Logged income is preferred to continuous income as it allows for a linear estimation, as the data would otherwise not be normally distributed. This means that the findings presented are interpreted as percentage changes in earnings. Whilst an additional attempt is made to contextualise this financially using the UK median income within the Chapter 3 discussion, a percentage change is still an easily interpretable measure.

The BCS70 data containing multiple observations of income at ages where income is reflective of maximum earnings is a strength. Time-averaged income is a better reflection of the individual's typical life-time earnings, allowing for a more general conclusion about the effects of language than would be feasible with income measured only at one age.

The limitation of the high retirement rate in the Newcastle Thousand Families Study was discussed at length within Chapter 2. It is possible that age 62 incomes would be more reflective of labour market earnings in future cohorts as the average retirement age increases, and effects within this age group may be present in later cohorts.

The cross-cohort approach leads to important differences in the measure. Household income is captured in Chapter 2, whereas individual income is captured in Chapters 3 to 5. The fact that both effects are significant means that two separate inferences can be made. Firstly, that the earnings that most strongly determine quality of life, one's household income, are significantly associated with SLCN. And secondly, that the earnings that most reflect the value of one's skills in the labour market, individual income, are significantly associated with receptive vocabulary (for which low scores have previously been used to identify SLCN as a sufficient, but not necessary, condition).

One final note on the occupational finding is that the effect is found after controlling for SES. Furthermore, the average levels of SES variables within high vocabulary groups do not significantly predict earnings after controlling for education level, cognitive and non-cognitive skills. Low language, or SLCN, limits the earnings potential of those identified relatively equally across the SES spectrum. Further research would be needed to quantify whether alternative methods would detect any moderation effects.

6.3.4.2 Education

Education has changed significantly in the UK since 1947 (Bolton, 2012). Therefore, education was primarily measured by the achievement of specific milestones. CASMIN levels were used in Chapters 3 to 5 to indicate levels of achievement that are relatively applicable across different education systems (Brauns et al., 2003).

The use of binary indicators is not an issue for linear estimation, although it does reduce the potential variability in academic achievement to a simplistic indication of success. Whether different subjects were studied is not captured by these measures, meaning that the larger effect of language disorders on reading, as opposed to maths (Young et al., 2002), is not observed.

Furthermore, the distributions of CASMIN levels are vastly different to modern expectations. Within the Newcastle Thousand Families Study, 37% of the sample had obtained no formal academic qualifications by age 50. This may reflect the labour market of the era, where academic qualifications may not have been necessary to start work in places where that would no longer be the case today. This is a limitation to the interpretation of education results, as we do not know how the SLCN group would have performed if they existed in the modern compulsory education system. However, the 11+ results and the odds ratios reported relating to all levels of education suggest that this SLCN group would be unlikely to succeed in any education system, relative to typically developing peers.

Chapter 4 demonstrated a clear endowment effect in relation to educational achievement. That is, higher vocabulary groups (multiple thresholds tested with consistent results) are more likely to obtain a degree, and this indicator in particular drives a sizeable portion of the observed earnings gap between high and low vocabulary groups. Furthermore, this impact was larger within the female-only sample. In the terms of the thought experiment carried out by the Oaxaca decomposition, if the low vocabulary females could achieve higher level education qualifications, this would theoretically correct around 40% percentage of the observed earnings gap.

6.3.4.3 Non-Cognitive Skills

The 60-item measure of the OCEAN non-cognitive skills (McCrae and Costa, 1987) in the Newcastle Thousand Families Study is an important strength. 12 items are recorded per characteristic, allowing a high level of reliability in the associations observed.

By comparison, the non-cognitive skills measured in the BCS70 are much more limited. The 15-item questionnaire measuring locus of control is relatively detailed, although locus of control represents only one subset of emotional stability. Furthermore, the teacher-reported measures of agreeableness, emotional stability and extraversion were not significantly associated with income, suggesting that they do not capture the traits detailed within Heckman and Kautz (2012). The conscientiousness items reported in Appendix Table A3.1 provide an example of the limited nature of the items, not intended originally to measure this characteristic.

Therefore, the conclusions drawn from this thesis relating to non-cognitive skills are focused largely on Chapter 2.

6.3.4.4 Health and Quality of Life

This thesis found generally equivalent levels of health and quality of life at age 62 for individuals with and without SLCN, regardless of the persistence or severity. However, there was a significantly higher psychiatric risk observed at age 50 amongst the SLCN sample, although this was not related to persistence or severity.

The general conclusions indicate that there is relatively little relationship between SLCN and health. Reliable measures of health were used, from the General Health Questionnaire (Goldberg and Williams, 1991) and grip strength measured at a clinical assessment. The CASP-19 scale used to measure quality of life is also considered a reliable measure within the 62-year old sample used (Hyde et al., 2003).

The results of this thesis suggest that the long-run economic costs associated with early speech, language and communication needs are not expected to be centred around later life healthcare costs, although this is only based on evidence at age 62. Furthermore, the individual's self-reported quality of life is not expected to significantly differ at age 62 given early SLCN. However, these findings do contradict the literature (Law et al., 2009; Le et al., 2020b), and the results presented are drawn entirely from one sample. It is possible that these effects reduce over time, as other factors grow in importance for impacting health and quality of life, and this is why the results at age 62, the latest age in the literature, are not significant.

6.3.5 Covariates

6.3.5.1 Socio-economic status

Socio-economic status was an important risk factor in the literature review for both communication and every later life outcome. The measurement of socio-economic status is highly variable across the reviewed literature, with a number of different proxies displaying similar effects when used in isolation. Maternal education tends to be the strongest indicator of socio-economic status where a choice exists (Reilly et al., 2010; Letts et al., 2013; Harrison and McLeod, 2010).

Within the Newcastle Thousand Families Study, the social class of the father's occupation was used, as this was the only available indicator. This was measured at birth and age five, with high levels of stability over this time allowing the inclusion of the birth measure alone for analysis.

The British Cohort Study contained a more reliable indication of socio-economic status, containing paternal occupation in addition to maternal and paternal education. This allowed for the analysis in Chapter 4, where background factors were used as additional predictors of the income gap between low, middle and high vocabulary groups. However, cognitive, non-cognitive and child education measures were also included in this analysis, limiting the potential for socio-economic status itself to have further impact.

The results of Chapter 4 demonstrate that socio-economic status does not have any additional impact on earnings in relation to language beyond those captured in cognitive and non-cognitive skills or education outcomes.

6.3.5.2 Gender

Gender assigned at birth, or sex, was measured in both datasets. Within Chapter 2, this was a significant predictor of SLCN status, with males around 1.75 times as likely to be identified with some form of SLCN. This was utilised as a control variable in order to recognise the important confounding effect that this variable would otherwise have in predicting the life outcomes measured.

Girls score higher than boys on average on the English Picture Vocabulary Test within the British Cohort Study. This has been previously found and discussed in Law et al. (2013), with explanations such as the receptive focus and the population sample hypothesised to explain the finding. The Oaxaca-Blinder Decomposition within Chapter 4 finds no significant endowment effect of higher proportions of males within the high vocabulary group when predicting income. Furthermore, Chapters 4 and 5 find no significant differences when restricted sampling is used to compare samples of males relative to females.

6.3.5.3 Cognition

Chapter 2 does not contain a measure of cognition in the first seven years of life. This is a limitation of any causal inference from results, as we cannot confirm that differences found in outcomes are a result of general cognitive deficits as opposed to speech and language.

Chapter 3 contains control variables relating to more general cognitive ability at ages five and ten. Visual motor co-ordination, reading ability and maths assessment scores are all included as control variables within the estimation of the association between receptive vocabulary and earnings. Chapter 4 extends this analysis by characterising these variables as cognitive covariates. Higher levels of the three cognitive covariates are found within higher vocabulary groups, explaining a significant portion of the observed income gap between vocabulary groups.

Three children are noted as having Special Educational Needs (SEN) within Chapter 2, although all three participants leave the study prior to the first outcome measure at age 11.

Within the British Cohort Study, we do not identify SLCN groups directly, using low language scores in various ways (continuous, arbitrarily grouped into thirds, and marginal thresholds). These groups are not identified methodologically as having language disorders, although this has been done previously through a strict threshold and a “sufficient but not necessary condition” argument (Law et al., 2009). Norbury et al. (2016) estimate that around a fifth of children with language disorders have other primary conditions, such as ADHD or learning disabilities. Non-verbal ability is used in alternative research papers (e.g. Law et al., 2009) to indicate this, whereas this thesis includes all individuals with SLCN as members of the population we seek to learn about. This is in line with the recommendations of Bishop et al. (2017), that this portion of people with SLCN should not be excluded from research. Approximately 2.7% of children at age 16 are identified with either a physical disability or a learning disability (Burchardt, 2004). However, whether the learning disability is primarily SLCN or another diagnosis is not reported. Given that Special Educational Needs were largely considered only in terms of severe disabilities prior to the Warnock Report of 1978 (Law et al., 2019b), it is unsurprising that there are no identifiers of SEN that can be related to SLCN in a way that would facilitate current policy recommendations. Therefore, Chapters 3 to 5 also cannot accurately distinguish the impacts of SEN from vocabulary or SLCN, and instead includes this small percentage of the sample without the use of this identifier.

6.3.6 Analytic Methods

Multiple analytic strategies are employed across the four analysis chapters. These methods allow for a broad range of outcomes and research questions to be addressed across the analysis chapters. Chapter 2 employs both linear and ordered logistic techniques. Chapter 3 uses mediation analysis via structural equation modelling, including the use of bootstrapping to construct confidence intervals. Chapter 4 utilises a Oaxaca-Blinder decomposition technique to assess counter-factual outcomes. Chapter 5 utilises a similar counter-factual technique of regression discontinuity design.

The broad methodological approach is a strength of this thesis. Without the wide range of methods, the number of research questions outlined and answered would be greatly reduced, or at least answered in a less robust way. Each analytic strategy is targeted at specific outcomes and research questions, all intended to maximise the strength and robustness of the conclusions that can be drawn.

Aside from some decisions regarding the SLCN identification from individualised records in Chapter 2, this thesis is primarily quantitative. This approach enables the wide range of questions and outcomes which can be reliably compared in their effects and allows for statistical tests of significance. However, qualitative approaches can offer greater detail in the underlying mechanisms explaining the observed associations. The conclusions of this thesis are drawn with an awareness of the limitations of the measures used and the exploratory nature of the initial research questions.

The conclusions of Chapter 3 regarding indirect effects through education could be considered to be the least reliable results within this thesis. These results are the only findings subject to change through a robustness check, where an alternative method for dealing with missing data resulted in significant mediation effects. There are various limitations, discussed within the chapter, to the measures used. As mediation essentially requires two significant linear relationships, this method is more susceptible than the other methodologies in the thesis to these limited measures, as they can impact multiple pathways. On reflection, two approaches may have been better; use of a dataset with more accurate and continuous measures, particularly of the mediating variables, although the author is not aware of any currently feasible alternatives (see section 1.3.5.2); applying the mediation logic to cognitive mediators and modelling a cumulative indirect effect only (i.e., does vocabulary predict income through its impact on other factors in childhood, adolescence, and early adulthood?).

6.3.6.1 Counter-Factual Approaches to dealing with Covariates

This thesis deals with the presence of covariates: variables which predict both verbal communication and outcomes, in three ways.

Chapters 2 and 3 treat the covariates in line with most of the quantitative literature, controlling for their impact using multivariate regression techniques. This allows for the estimation of the impact of the verbal communication measure whilst holding the covariates equal. However, this approach is limited, and causal inference cannot be not drawn from these analyses. We are only able to account for the observed covariates, whilst unobserved factors may bias our estimates if they causally impact both our independent and dependent variables. Endogeneity of this form has been a cornerstone of the econometrics literature for decades, informing the selection of alternative modelling methods for Chapters 4 and 5 (Angrist and Pischke, 2009).

Chapter 4 attempts to highlight the importance of the observed covariates by conducting a counter-factual thought experiment. This analysis quantitatively hypothesises the outcomes in alternative worlds where low vocabulary individuals are not similarly endowed with low levels of valuable covariates (such as SES, cognitive and non-cognitive skills or education outcomes). This is an alternative approach to considering the “other things equal” method of estimating the importance of language drawn from the econometrics literature (Angrist and Pischke, 2015; Oaxaca, 1973; Blinder, 1973).

The use of the decomposition in Chapter 4 helps deal with the limitations of a linear mediation model in Chapter 3. The relationship between receptive vocabulary at age 5 and earnings in adulthood is continuous and cumulative. However, it is not a large effect observable over small movements in the vocabulary score (see Chapter 5). Therefore, a linear mediation model may have been too sensitive to pick up the underlying mechanisms using conventional statistical significance. Instead of modelling “Does an increase in early vocabulary predict a change in binary measures of education and subsequently predict an increase in annual income?”, as in Chapter 3, Chapter 4 simply compares the average earnings of two groups and statistically attributes the observed differences (which are large across distinct parts of the distribution) to the range of explanatory variables which also differ between groups.

In summary, Chapter 4 is more relevant when trying to understand the key differences between children with high and low vocabulary, whereas Chapter 3 is more relevant when questioning the potential impact of a gradual increase in early vocabulary.

The evidence of Chapters 2 to 4 demonstrates the importance of these covariates in determining outcomes for individuals with different verbal communication skills. Within the data collected, it is difficult to make clear pathway models due to the limited number of time-points and measures, as evidence by the non-significant mediation effects in Chapter 3. Therefore, Chapter 5 takes an alternative counter-factual approach to dealing with the covariates. The Regression Discontinuity Design (RDD) approach equalises the covariates between a control and a treatment group by utilising random variation around a narrow threshold. This is considered to be close to a Randomised Control Trial (RCT), often considered the gold standard method for causal inference, as the random assignment in RCTs deals with the impact of all observed and unobserved covariates (Schulz et al. ,2010).

The RDD approach within Chapter 5, however, has a benefit over RCT methods employed where participation is voluntary. Within Chapter 5, participants would be unaware of the assignment taken in the analysis, meaning that the randomisation samples remain equivalent after assignment. In some RCT trials, voluntary participation may cause differences between the treatment and control groups after the assignment (Angrist and Pischke, 2009). Whilst the analysis is only valid at a narrow threshold, the inferences made about the value of thresholds for predicting outcomes can be interpreted as causal, as they are not influenced by any observed or unobserved covariates within this range.

The following recommendations for future research reflect experience drawn from the experience of conducting this research, but also a summary of potentially interesting future avenues opened to uncover the intricacies behind associations found within this thesis which may be appropriately tackled through quantitative or qualitative methods.

6.4 Recommendations

The current thesis examines the outcomes for two valuable cohorts. The results of this thesis should be interpreted in light of the historical context. No interventions were analysed, and differences between children's development of verbal communication cannot be inferred as causal producers of the outcome differences observed.

The recommendations below represent a combination of conclusions derived from the research questions answered and conclusions derived from the review of the literature.

6.4.1 Policy

This thesis provides evidence on important questions relating to a recognised group in society. The conclusions drawn from these analyses contain important messages with clear policy relevance.

6.4.1.1 Education

Firstly, this thesis supports a direct policy focus on language in early years. Early communication predicts educational attainment at all levels within both cohorts. The education system changed dramatically across the two cohorts, with comprehensive schooling becoming the majority, replacing the tripartite system (Bolton, 2012). Whilst children with SLCN were significantly less likely to be accepted into a grammar or technical school than typically developing peers, neither schooling system was able to eradicate the existence of a substantial, consistent attainment gap from school entry through to further education.

Chapter 4 shows that changing the educational endowments of individuals with low vocabulary would diminish the earnings gap to high vocabulary individuals. This finding was particularly true for the low likelihood of low vocabulary females obtaining degree-level qualifications.

Resources targeted at funded early years settings or the home environment which support language development may help to reduce the number of children with lower life chances. Furthermore, approaches targeted within the school setting can provide support to a greater number of children, as very few receive external support for their speech, language and communication needs (McGregor, 2020; Norbury et al., 2016).

6.4.1.2 Service Provision

There is substantial evidence of the effectiveness of interventions for children with low language, whether associated with SES (Greenwood et al., 2020), or relating to Developmental Language Disorder (Law et al., 2017c; Law et al., 2017a).

This thesis does not model intervention effects or the causal effect of early language ability. However, this thesis does provide evidence of a large range of negative associations related to SLCN, and evidence of significantly lower earnings across adulthood for individuals with low language.

Therefore, this thesis argues that the short-term costs associated with provision of services for children with low language, or broader SLCN, are potentially far lower than any long-term benefits. Children with higher language skills tend to earn more, implying higher productivity and tax returns.

Snow (2009) state that the field of speech and language therapy has prioritised a focus on the individual with a disorder over the broader social population factors driving early communication. The importance of covariates and risk ratios demonstrated across this thesis support the use of targeted interventions, whilst the continuous modelling supports the use of universal support. Therefore, the appropriate recommendation is for proportionate universalism, whereby more resources are funnelled into disadvantaged areas to identify those at highest risk as early as possible. 'New' public health models (Baum 2002) concentrate on preventable factors such as economic well-being, lifestyle (including family stress), and environment (including family stress and resource as well as access to services). Preventative interventions are typically cheaper than remedial interventions (Law et al., 2013), although leveraging change across broader social determinants requires a substantial population-based effort.

Crucially, language is determined in part by status in early life, before subsequently driving later status in life (Baum, 2002; Law et al., 2013). The results of this thesis fully support both parts of this statement, as shown by the controlling impact of covariates and the broadly different outcomes for children with different early communication skills.

6.4.1.3 Identification of SLCN

Secondly, this thesis recommends a more detailed approach to assessing the validity of language assessments in diagnosing language disorders. Given the funding restrictions outlined in the Bercow Report (2008) and Bercow: Ten Years On (2018), efficient diagnosis and the targeting of resources to those most in need are crucial. This thesis represents the first analysis of the accuracy of a threshold value at the margin for detecting language disorders, with prior research measuring sensitivity and specificity as indicators of accuracy across the whole distribution of scores (e.g. Letts et al., 2013; Spaulding et al., 2006).

The core recommendation drawn from this chapter is that thresholds should not be applied strictly in policy documents, as they do not accurately identify the presence of functional difficulties. Children either side of any threshold display similar outcomes in many areas, implying similar capabilities. Whilst strict thresholds enable resource allocation, there is a high risk of misdiagnosis either side of the threshold. In particular, the risk of underdiagnosis is likely to be high given that fewer children access services than are estimated to suffer with SLCN (see McGregor, 2020 for a summary relating to Developmental Language Disorder). Chapters 2 and 3 demonstrate the large, persistent economic and personal costs associated with early communication difficulties. Missing the opportunity to provide additional support to the worst affected children through the application of poorly constructed threshold values (i.e., the thresholds are drawn at 1, 1.5 and 2 standard deviations below the mean through no specific theory relating to oral communication or SLCN to the author's knowledge) could be damaging to individuals and the economy.

Where different policies are applied either side of the threshold, an effective intervention applied to the individuals below that threshold may create an inequality at the threshold where none previously existed. Whilst these inequalities would be randomly distributed (as a result of children being unable to self-select their performance accurately around the threshold), as opposed to a social inequality, any inequality between children is undesirable where it could be easily avoided.

When measuring the accuracy of a language assessment, a sample of children with and without identified language disorders sit the assessment. To test the accuracy of the New Reynell Developmental Language Scales (a test measuring comprehension and production between ages three to seven (Edwards et al., 2011)), for example, Letts et al. (2013) test thresholds at 1.5 and 2 standard deviations below the mean for their accuracy of identification within the sample. Using the data of who has a disorder and the language assessment score, a regression discontinuity design could be easily employed to provide an additional measure of accuracy. The accuracy of a chosen threshold at the margin would provide an indication of whether this threshold can be strictly applied, as opposed to providing a general indication. This method is recommended as one potential way to increase the evidence base regarding the accuracy of language assessments. The value of this recommendation, however, is dependent upon the use of evidence-based practice in research and policy, considered by Nitido and Plante (2020) and Spaulding et al. (2006) to be relatively rare.

These results support the argument that strict threshold values should not be indicated in policy documents. However, a child's score on a language assessment does provide a general indication of ability and these scores are useful tools for identifying areas of strengths and weaknesses. The results of this thesis speak only to narrow movements around a threshold as a comment on the strength of priority that should be given to different criteria for being allocated funding for intervention.

6.4.2 Future Research

This thesis presents a wide array of results and approaches, many of which represent the first estimations of their kind in the field. This gives a wide scope for future research potential to develop out of these findings, some of which are outlined below.

6.4.2.1 Samples

This thesis highlights the value of existing cohorts for addressing important research questions. Whilst there are opportunities afforded by the detail in more recent cohorts, there are huge benefits to utilising older datasets which contain measures into adulthood. Early life skills are hugely important to life-course studies, and the reverse should be recognised. Outcomes in later life represent a huge source of interest from parents of children with SLCN (Dockrell et al., 2014) and this should be reflected in the breadth of research conducted.

6.4.2.2 Education

Chapters 2, 3 and 4 support the literature reviewed in Chapter 1, that educational attainment throughout school and further education is associated with the early development of speech and language. The literature reviewed implies that earnings are reduced through this association (Marsh et al., 2010), although the results of this thesis do not universally support this conclusion.

The Oaxaca-Blinder decomposition analysis suggests that individuals with low vocabulary levels are not poorly rewarded for the education levels they do achieve. By extension, improvements in the educational attainment of this group would be expected to be valued at the rates observed in the population-level literature. However, the mediation analysis did not find a significant mediation effect of education on earnings when vocabulary is measured continuously. This contradiction is likely to be a result of the poor predictive power of the model employed for estimating earnings. Further research would be beneficial to establish whether a focus on further education access represents a valid route to improving the earnings potential of individuals with a limited early receptive vocabulary.

The Oaxaca-Blinder decomposition revealed particularly strong degree-level endowment effects within the female-only sample. This suggests that girls with low vocabulary in this sample were highly unlikely to go to university, relative to girls with high vocabulary, and that this predicts a significant portion of the income gap between these groups. Females with low language are less likely to be referred to speech and language therapy services than males according to much of the literature reviewed in Section 1.2.5.2. The results of Chapter 4 indicate that low vocabulary females are relatively more disadvantaged than low vocabulary males (they have a larger earnings gap to high vocabulary individuals of their gender), and this may be partially explained by factors relating to support in the school setting. Alternatively, this could reflect the number of available routes for unqualified males, relative to females, in the labour market. Verbal communication may therefore play a role in explaining factors behind the broader gender pay gap, particularly at the lower end of the earnings distribution.

6.4.2.3 Occupational

Income within the British Cohort Study was clearly related to childhood receptive vocabulary. Cohorts such as ALSPAC (Avon Longitudinal Study of Parents and Children), born in 1991/92, are approaching an age where analysis of income would be broadly representative of career income. The expansion of research into alternative cohorts with greater detail in the measures of language, and income data which reflects the modern labour market, could build upon the foundational results of this thesis.

Furthermore, Chapter 3 is a product of the conclusions of Chapter 2. Associations were found with both education (including further education) and non-cognitive skills, both of which were novel findings at the time. This informed the selection of mediation pathways in Chapter 3. The recommendations below represent a combination of conclusions derived from the research questions answered and conclusions derived from the review of the literature. Future research could take a broader approach and consider any mutable factors occurring after the measure of language (or identification of SLCN in a more thorough dataset) as mediators, including education and non-cognitive skills alongside other cognitive measures. With more regular intervals of observation, and a sample with lower rates of attrition, more interactions could be estimated to build narrative conclusions about the overall impact of early language development.

Occupational outcomes more broadly would be of greater interest in cohorts which reflect the modern labour market. Within this thesis, earnings are used as an indicator of occupational success, yet this is only one element of the working life experiences of an individual. Earnings reflect only the value of the skills used within the job worked, meaning that individuals with

SLCN or poor receptive (or global) vocabulary may work in occupations which do not require high levels of verbal communication. Given that individuals with language disorders do not lack prosociality (Conti-Ramsden et al., 2013), this would reflect a source of disadvantage not captured in the simple earnings measure employed in this thesis. Further research into the specifics of occupational sectors and required skills could illuminate some of the mechanisms behind the earnings gap beyond the analysis conducted in Chapters 3 and 4.

6.4.2.4 Non-Cognitive Skills

This thesis finds the first associations with adult non-cognitive skills and early speech and language development. These skills develop throughout life and can be enhanced in several ways, as outlined in Chapter 1. Further research could establish the age at which important non-cognitive skills begin to diverge for children with communication difficulties. The timing of the non-cognitive skills measure in the Newcastle Thousand Families Study, and the lack of detailed measures in the British Cohort Study, limited the potential for addressing this relationship with the available data.

The most important non-cognitive skill in low SES populations has been argued to be openness to experience (Falk et al., 2016). This thesis finds that an SLCN population have lower openness to experience at age 62. A large facet of openness to experience is the enjoyment of educational opportunities and experience, the general joy of learning new things. The finding that this is significantly lower within a 62-year old group of individuals who had SLCN between the ages of two to seven could be a crucial contribution to future research.

Openness to experience is fostered from very early in life and only stabilises in adulthood (Heckman and Kautz, 2012). The finding that educational experiences are substantially worse for children who struggle to communicate is well established (Conti-Ramsden et al., 2013; Byers et al., 2012) and intuitive. It is therefore unsurprising that their educational attainment tends to be significantly poorer (Durkin et al., 2009; Conti-Ramsden et al., 2018). It is worth considering whether an individual's enjoyment of education is impacted by these experiences to such an extent as to permanently lower their openness to new experiences more generally.

Further analysis is necessary of the connections between the enjoyment of education, educational attainment, and early communication. Whilst receptive language difficulties are often considered to impact educational attainment through a poorer understanding of the curriculum and teacher's instructions (Hendricks et al., 2019), there may be a further effect occurring through the enjoyment of time spent in school. This would have implications for the approaches of interventions aimed at improving educational attainment. Helping students with communication difficulties to build relationships, as well as providing enjoyable experiences and reducing negative social experiences in school settings, could positively reinforce the enjoyment of education, generating a non-cognitive skill which is valuable throughout life.

Whilst openness to experience is crucial in low SES populations, due to the connection with education, conscientiousness is more commonly recognised as the most predictive non-cognitive skill for health and occupational outcomes. Conscientiousness reflects work ethic, attention to detail, and attention span amongst other traits. Attention span is commonly linked to SLCN in early childhood (Westrupp et al., 2020; Beitchman et al., 1996; Harrison and McLeod, 2010; Levickis et al., 2018a), yet levels of conscientiousness are statistically equivalent at age 62 for SLCN and typical individuals within Chapter 2. Conscientiousness could represent a point of strength to emphasise for this population, although this would require evidence of the age at which these differences may be expected to reduce. Conscientiousness was not predicted by vocabulary within Chapter 3 in a multi-variate regression, and the slight differences in mean levels of conscientiousness across vocabulary groups did not explain any of the income gap in Chapter 4.

6.4.2.5 Socio-Economic Status

Future studies may seek to reveal more detail about the potential interactions between socio-economic status and verbal communication. Higher socio-economic status may influence the access to services, through parental or neighbourhood factors (Longfield et al., 2019), representing a potential protective factor.

Alternatively, lower socio-economic status may be characterised as a risk factor for exacerbated effects of SLCN. This would depend on whether communication is a vital skill for social mobility, as is implied by the Department for Education's (2017) priority in closing the socio-economic word gap.

6.4.2.6 Analytic Techniques

Application of the analytic techniques presented may also provide opportunities for future research. The Oaxaca-Blinder decomposition is suitable to any comparison of two groups on any continuous outcome. For example, when analysing the relationship between internalising and externalising Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997) scores and language, the impact of gender is commonly noted (e.g. Forrest et al., 2018). A formal comparison of males and females using this methodology could help to describe some of the key differences, as it explicitly separates the gender differences in initial language scores from the gender differences in the relationship between language and the SDQ scores.

6.4.2.7 Identification of SLCN

Chapter 5 highlights and discusses the value of thresholds on language assessments. In Section 6.4.1.3, it is recommended that language assessments should model the accuracy of recommended thresholds at the margin. The recommendation for researchers using thresholds is comparatively simple and easily achievable - researchers should report the reasoning behind a threshold choice.

Thresholds drawn further from the mean will result in identification of a more severely impacted group and this fact should be retained in the description of any conclusions drawn. Moreover, the sensitivity of results to changes in the chosen threshold are computationally simple to calculate and can increase the reliability associated with the main results reported.

6.5 Conclusions and Final Remarks

A substantial portion of society experience difficulties with communication in childhood. The children with speech, language and communication needs (SLCN) may resolve these difficulties early in life, or they may persist at least into schooling.

Experiences of education are clearly poorer for those with poorer communication skills. This thesis also supports the conclusions of Conti-Ramsden et al. (2018) that the attainment gap in education for those with SLCN impact the individual’s probability of achieving qualifications at all levels of further education. These findings are particularly strong for those with persistent or severe difficulties in both cohorts examined.

Inclusion of non-cognitive skills in the Newcastle Thousand Families Study revealed associations between SLCN and two of the OCEAN characteristics – openness to experience and extraversion.

The association with openness to experience is reported at age 62, although the noted attainment statistics across all years of education imply that this may be established far earlier. Further research within modern cohorts may reveal a potential pathway to improving educational attainment through the targeting of this non-cognitive skill, or at least increase understanding of the mechanisms that drive the educational attainment gap.

Lower extraversion within this group is consistent with the expectations drawn from the literature. Social confidence is an understandable area of relative weakness for those who struggle to communicate, and this remains an important area of research. The finding that this persists into their overall levels of extraversion at age 62 highlights the life-long impact of the development of communication skills in the early years of life.

Three personality traits, by comparison, were not significantly associated with the identification of SLCN. Emphasis and enhancement of areas of strength are an important element in supporting the outcomes for disadvantaged groups. Conscientiousness, commonly considered the most important of the OCEAN skills, is not significantly impacted by early communication within this sample. Agreeableness was also found not to be associated with SLCN within this sample. The literature relating aggression in early years to SLCN does not persist throughout life according to these results. People with a history of SLCN displayed equal levels of warmth and friendliness in adulthood, as is consistent with literature regarding their prosociality from childhood (Toseeb et al., 2020).

Health and quality of life outcomes in the Newcastle Thousand Families Study showed almost no association with SLCN. This is quite contradictory to the prior literature. Psychiatric risk, at age 50, was related to SLCN status in Chapter 2, consistent with the association at age 34 in Law et al. (2009). However, this effect did not persist to age 62, where measures of objective health and quality of life were also comparable between SLCN and typical groups. Anxiety has precedence in the literature for association with SLCN (Durkin and Conti-Ramsden, 2010), yet levels of emotional stability were also equal within this sample at age 62. Whether these contradictions of the literature are driven by the length of time studied, reducing the proximity of effect, or through other factors relating to the samples used require further research.

Analysis of the British Cohort Study supports the suggestive conclusions within the Newcastle Thousand Families Study, that earnings in adulthood are significantly reduced for those with lower language ability. This has implications for discussions around funding. If increased support for language development in childhood (whether in schools, home environments or speech and language therapy services) results in greater productivity across adulthood, the long-term benefits of this initial support could be substantially greater than the initial cost.

This thesis demonstrates the value of language assessment scores. These scores provide a clear indication of the ability to communicate and can predict the later earnings of the individual three decades later from this. However, these scores are only an indication and not an accurate ranking. Whilst this is generally true of any assessment given the existence of measurement error, the limits of language assessments should be considered in their application. When applying threshold values, as opposed to a continuous score, the value chosen will always impact the findings (Nitido and Plante, 2020). The modelling approach in Chapter 5 highlights that commonly used thresholds are no more reliable than any other threshold. Furthermore, no threshold was able to accurately separate individuals at the margin on the basis of their language ability having functional impacts on their life.

Life is by no means predictable, either statistically or anecdotally, yet there remains an association between early communication and later life chances in many areas. The development of speech and language in early years clearly matters and supporting children to develop these important skills may impart greater benefits over the life course than were previously realised.

Abbreviations

BCS70	British Cohort Study 1970
CASMIN	Comparative Analysis of Social Mobility in Industrial Nations (Education Levels)
DLD	Developmental Language Disorder
GHQ	General Health Questionnaire
GLS	Generalised Least Squares
GSEM	Generalised Structural Equation Modelling
OCEAN	Big 5 Personality Characteristics: **O**penness to experience, **C**onscientiousness, **E**xtraversion, **A**greeableness, emotional stability (- **N**euroticism)
OLS	Ordinary Least Squares
RCSLT	Royal College of Speech and Language Therapists
SES	Socio-Economic Status
SLCN	Speech, Language and Communication Needs
SLI	Specific Language Impairment (N-SLI = Non-Specific Language Impairment)

Appendices

Chapter 2 Appendices

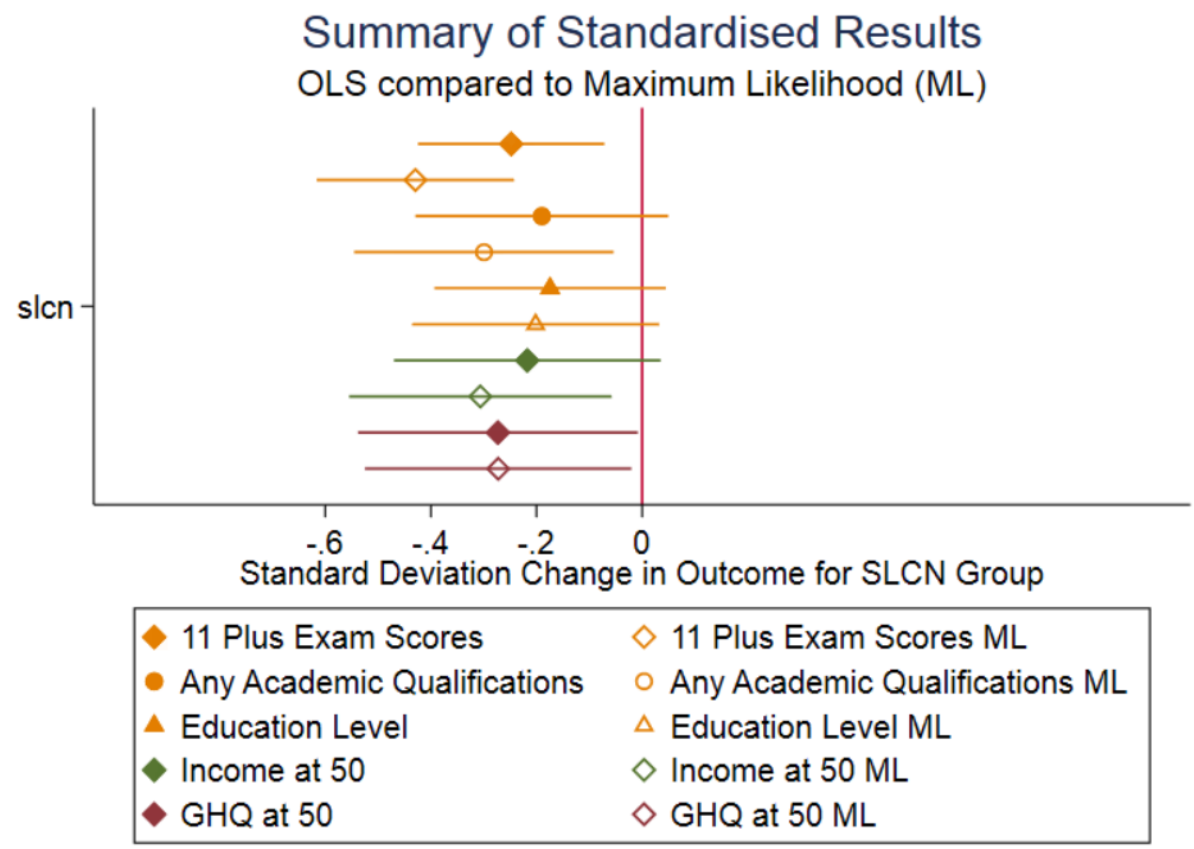
Appendix Table A2- 1 Prevalence of SLCN by gender and school type

		Secondary Modern	Technical School	Grammar School	Total
Full Sample	**'Typical'**	432	60	72	**564**
	SLCN	115	13	7	**135**
Males	**'Typical'**	190	43	36	**269**
	SLCN	72	10	5	**87**
Females	**'Typical'**	242	17	36	**295**
	SLCN	43	3	2	**48**

Appendix Table A3- 1

Males are less likely to display 'Typical' speech and language development. There is no gender difference in attendance of grammar or secondary moderns although males were significantly more likely to attend a technical school than females.

Typically developing children of both genders were significantly more likely to attend a grammar school.



Appendix Figure A2- 1 Maximum likelihood estimation results for converging models

Note: Comparison of equivalent regressions estimated using OLS results, as presented in the main text, and the primary alternative methodology of Maximum Likelihood Estimation. Only five outcomes converged in the ML model.

Chapter 3 Appendices

Appendix Table A3 1- Summary Statistics for Control Variables (three parts)

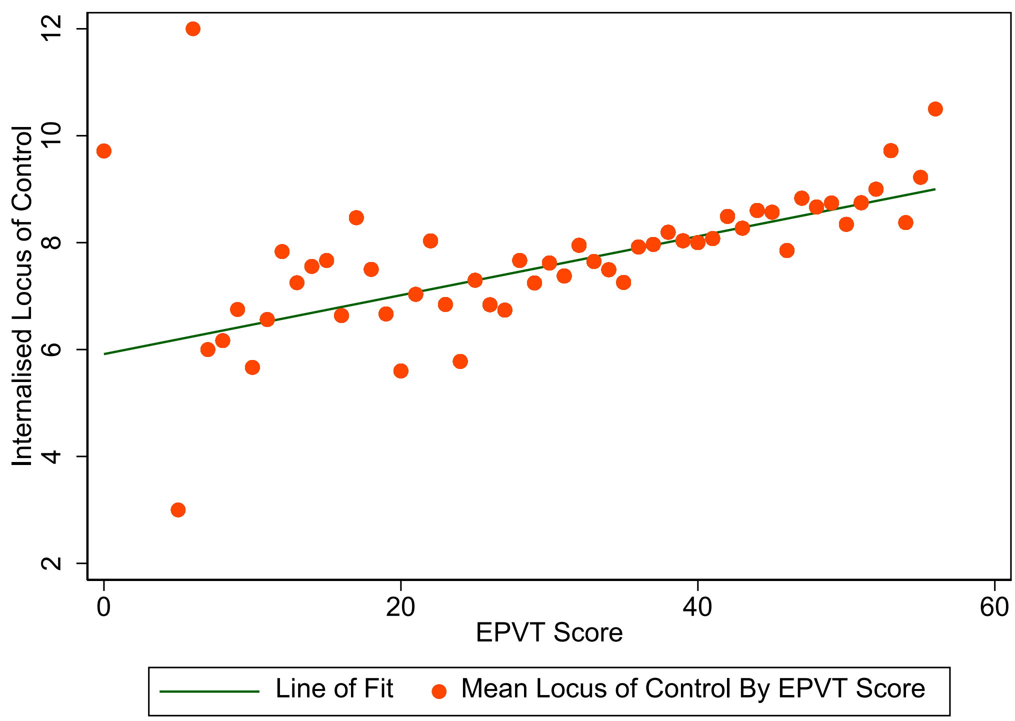
Variable	Mean	Std. Dev.	Min	Max
Copying Designs	5.15	1.94	0	8
Friendly Maths Test	47.66	11.34	3	71
Edinburgh Reading Test Shortened (standardised by CLS on full sample)	0.27	0.93	3.08	1.95

Variable	*Group within Variable*	*N*	*%*
Gender	*Male*	*1306*	*50.5*
	Female	*1282*	*49.5*
Father's Education	*CASMIN Level 1*	*1345*	*52.0*
	CASMIN Level 2	*731*	*28.3*
	CASMIN Level 3	*512*	*19.8*
Mother's Education	*CASMIN Level 1*	*1648*	*63.7*
	CASMIN Level 2	*677*	*26.2*
	CASMIN Level 3	*263*	*10.2*
Father's Occupation (Mother's if missing) by Registrar General Social Class + Unemployed group	*Unemployed*	*91*	*3.5*
	Unskilled	*95*	*3.7*
	Partly Skilled	*318*	*12.3*
	Skilled Non-Manual	*1167*	*45.1*
	Skilled Manual	*389*	*15.0*
	Technical / Managerial	*343*	*13.3*
	Professional	*185*	*7.2*

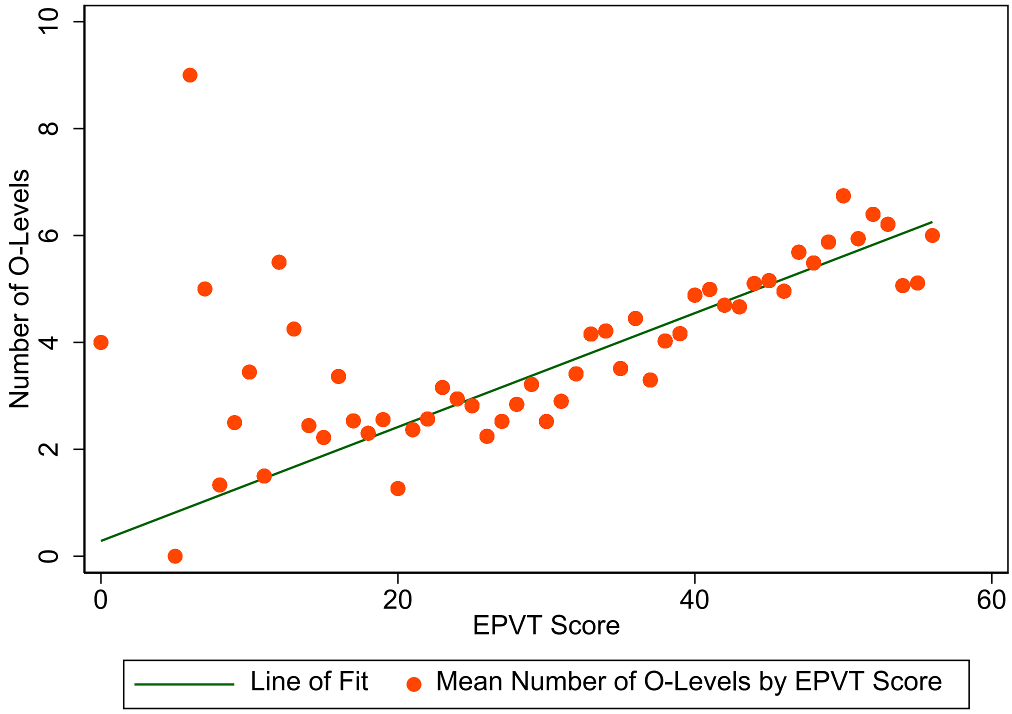
Appendix Table A3.1 continued

Facet	*Description*	*BCS Item*	*Mean*	*Std. Dev*	*Min*	*Max*
Impulse Control	*"The propensity to inhibit proportional responses"*	*Easily Distracted*	*21.02*	*22.94*	*1*	*100*
Orderliness	*"The propensity to be organised and neat"*	*Difficulty Concentrating*	*29.88*	*28.15*	*1*	*100*
Decisiveness	*"The propensity to make deliberate decisions"*	*Cannot Settle*	*25.38*	*24.85*	*1*	*100*
Reliability	*"The propensity to work hard"*	*Fails to Finish Things*	*19.54*	*21.52*	*1*	*99*
Total Conscientiousness (400 minus summation of scores)			*305.18*	*79.58*	*7*	*397*

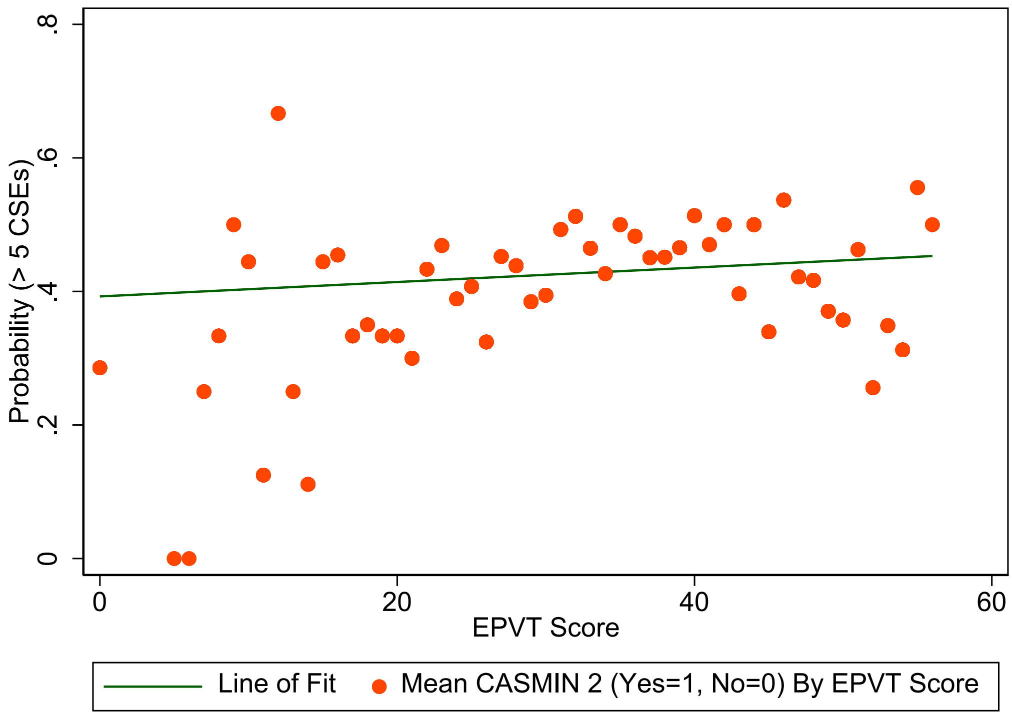
Note: Facets, descriptions and items from Prevoo and ter Weel (2012) with summary statistics taken from the final regression sample used in the current analysis (n=2588).



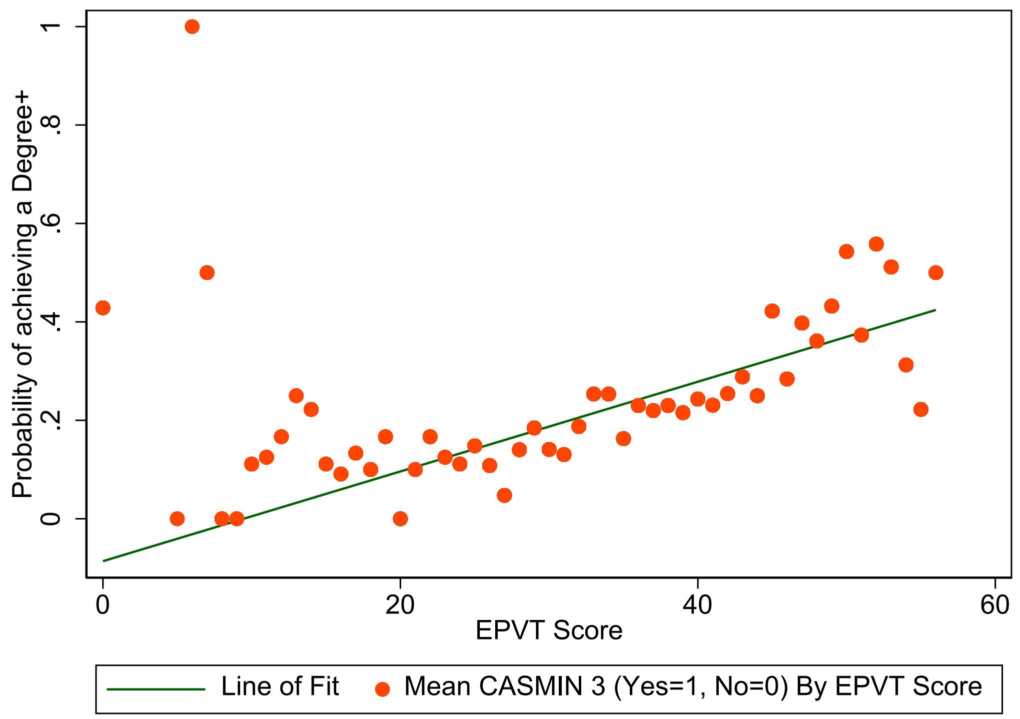
Appendix Figure A3 1- Trend for Mean Locus of Control (0-15) against EPVT



Appendix Figure A3 2- Trend for Mean Number of O-Levels (0-15) against EPVT



Appendix Figure A3 3- Trend for attaining CASMIN Level 2 (probability) against EPVT



Appendix Figure A3 4- Trend for attaining CASMIN Level 3 (probability) against EPVT

Appendix Table A3 2- Alternative Bootstrapped Confidence Intervals Estimating the Direct Effect of EPVT Scores on Income

	Percentile Confidence Interval		Bias-Corrected Confidence Interval		Bias and Acceleration-Corrected Confidence Interval	
Direct Effect	Min CI	Max CI	Min CI	Max CI	Min CI	Max CI
	0.0002	0.0056	0.0003	0.0057	0.0003	0.0057

Note: A 0.001 size effect would suggest a gain of 1 word on the test correlates to a 0.1% increase in annual income. Approximately 10 times this effect relates to a gain of 1 standard deviation on the EPVT (Std. Dev. = 9.86).

The above direct effects are estimated using standard, bias-corrected (proportion of bootstrapped estimates below the effect estimate) and bias plus acceleration-corrected (skew in the distribution of bootstrapped estimates). Consistent results are reported within all three methods, supporting the results within the main text.

Appendix Table A3 3- Results of Model Specification without Number of O-Levels (three first Stage regressions and reduced final stage)

	Direct Receptive Vocabulary	Indirect through Locus of Control	Indirect through CASMIN Level 2	Indirect through CASMIN Level 3	Indirect through Number of O-Levels	Total Indirect Through Education
Effect on Time-Averaged Income	0.0026** *(0.0012)*	0.00005 *(0.00004)*	0.00005 *(0.00012)*	0.00007 *(0.00028)*	N/A	0.00011 *(0.0002)*

p<0.1 = *, p<0.05 = **, p<0.01 = ***

Note: Bootstrapped (10,000 replications) standard errors in parentheses

Appendix Table A3 4 Results of Sobel Test of significance for full model specification.

	Direct Receptive Vocabulary	Indirect through Locus of Control	Indirect through CASMIN Level 2	Indirect through CASMIN Level 3	Indirect through Number of O-Levels	Total Indirect Through Education
Effect on Time-Averaged Income	0.0029** *(0.0013)*	0.00005 *(0.00005)*	0.00001 *(0.00005)*	0.00011 *(0.00025)*	0.00015 *(0.00014)*	0.00027 *(0.00032)*

p<0.1 = *, p<0.05 = **, p<0.01 = ***

Note: Sobel standard errors in parentheses

The Sobel (1982) test compiles a joint significance z estimate for the indirect pathway as an alternative significance test to the bootstrapping method. Our results are consistent with this method of analysis, as shown above in Appendix Table A3.4.

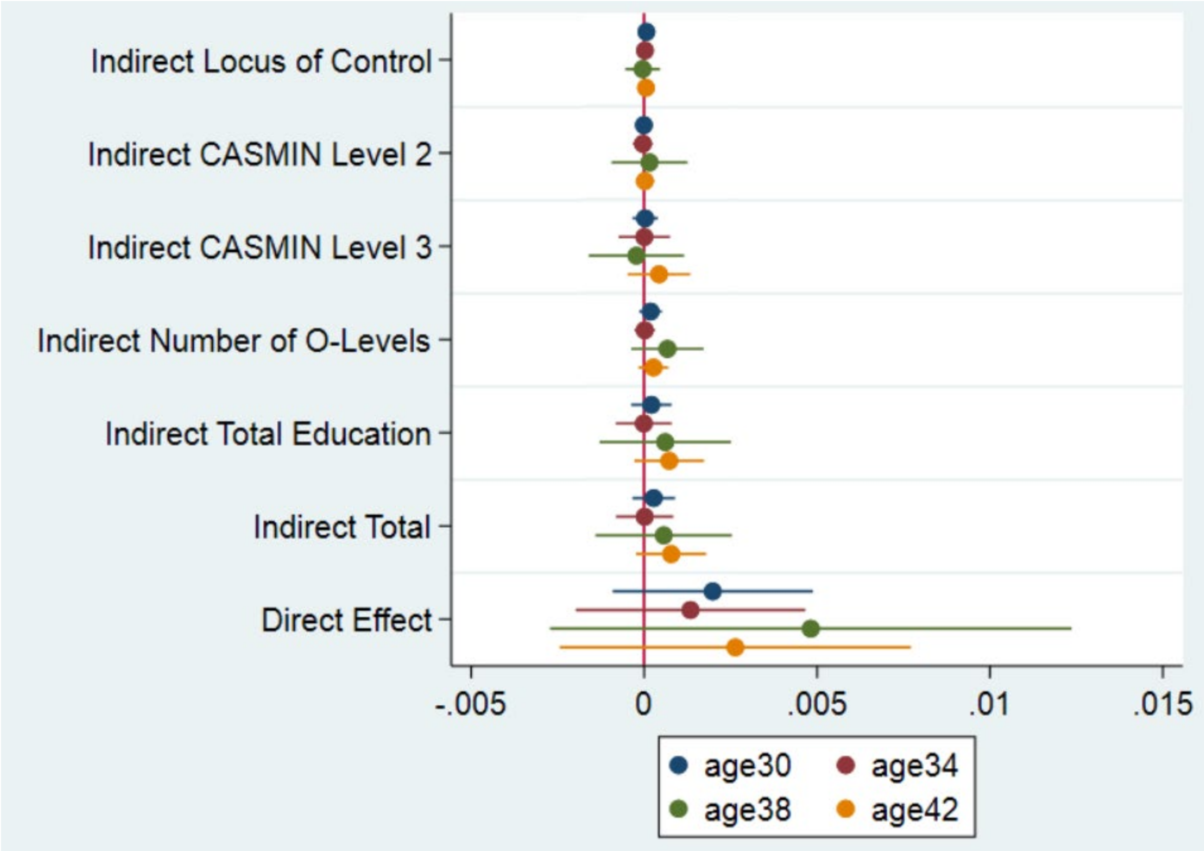
Appendix Table A3 5- Results of Baron and Kenny Tests of individual δ significance for full model specification (Full information sample each time, n=2588).

	Outcome	Locus of Control (δ_1)	CASMIN Level 2 (δ_2)	CASMIN Level 3 (δ_3)	Number of O-Levels (δ_4)
Stage 1 Regression	Effect of EPVT Score *(Standard Error)*	0.0082 *(0.0055)*	0.0003 *(0.0011)*	0.0004 *(0.0008)*	0.0071 *(0.0067)*
		Locus of Control (δ_5)	CASMIN Level 2 (δ_6)	CASMIN Level 3 (δ_7)	Number of O-Levels (δ_8)
Stage 2 Regression	Effect on Time-Averaged Logged Income *(Standard Error)*	0.0063 *(0.0046)*	0.0400 *(0.0375)*	0.297*** *(0.052)*	0.020*** *(0.005)*

p<0.1 = *, p<0.05 = **, p<0.01 = ***

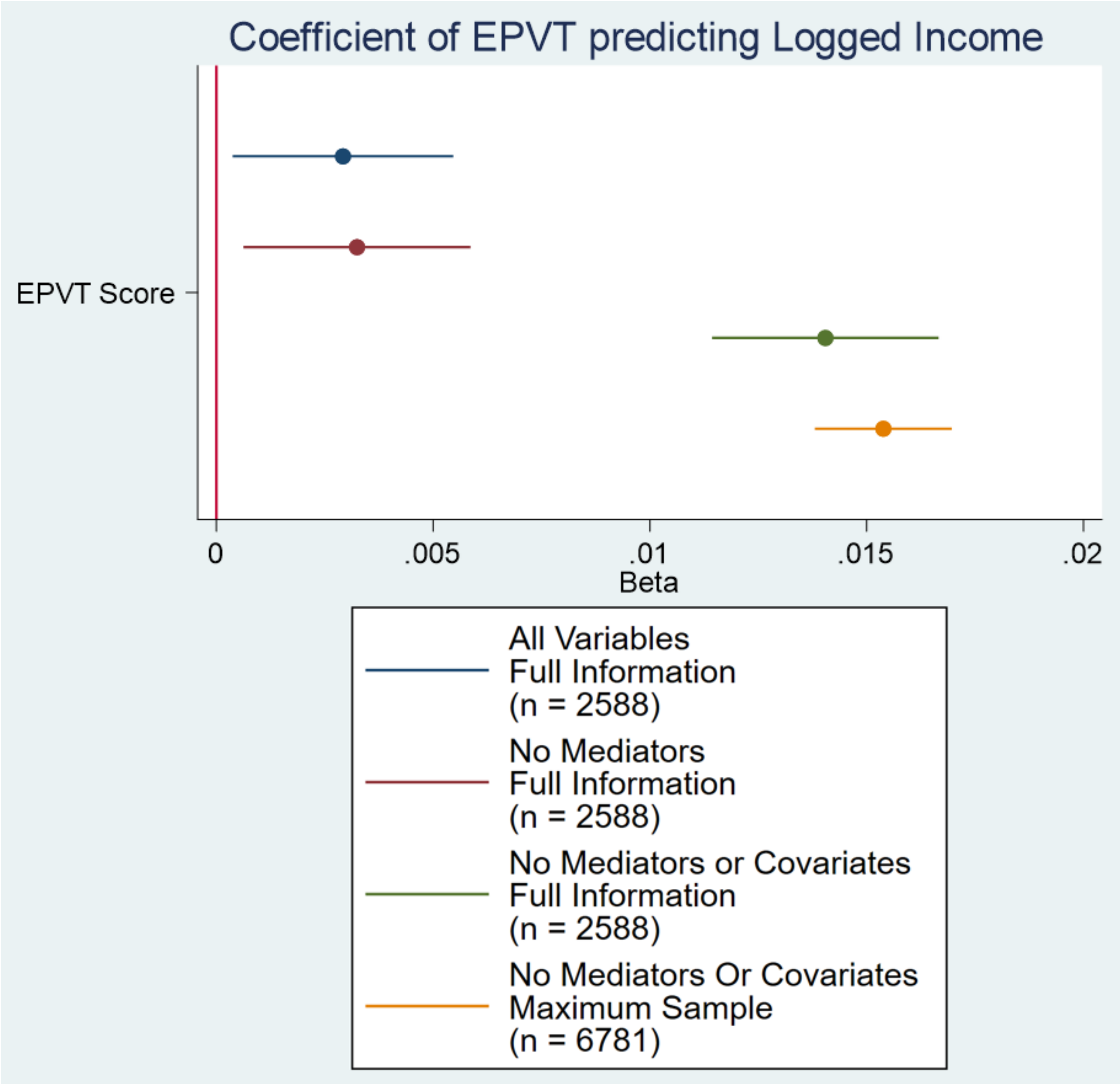
Note: OLS Standard Errors in parentheses

The original Baron and Kenny recommendation was to measure significance of an indirect pathway via the criteria that each of the two pathways were themselves significant. As a more conservative method, this too is consistent with our above finding that the mediation indirect pathways are not significant. As reported in Appendix Table 3.5 only δ_7 & δ_8 are individually significant and the other 6 δs are not.



Appendix Figure A3 5- Coefficients derived from the model specification ran separately at each time point income is observed to compare

*Note: A 0.01 effect would suggest that a 1 word increase on the language test correlates to a 1% increase in income in that year. Circles denote point estimate with a 95% confidence interval (1.96*std. error either side) denoted by the horizontal lines. If the lines did not overlap, this would suggest a significantly different effect between time points for a given effect. They do overlap on all time points for each estimate showing that the time of income measurement does not significantly alter our estimates.*



Appendix Figure A3 6- The direct effect estimate of a 1 word increase in the EPVT score on logged incomes, plotting the point estimate and 95% Confidence Interval.

This figure compares the coefficients obtained from four models. Firstly, the model reported in the main text is shown. This is broadly similar to the exact same model and sample but without the mediator variables. However, when the covariates are removed from the model (green line, model three), the coefficient changes dramatically. Model four repeats model three but using the full sample available when only using the two remaining variables (income and EPVT score), as many individuals had been initially removed due to missing data on the mediators and covariates.

What this highlights is that sample selection does not drive the largest difference in modelled effect size, but covariate inclusion does. In short, failing to account for the controlling effect of covariates multiplies the observed effect by almost five.

Appendix Table A3 6- Results of model specification (removing female dummy variable from controls) ran only for male sample (n=1306)

	Direct Receptive Vocabulary	Indirect through Locus of Control	Indirect through CASMIN Level 2	Indirect through CASMIN Level 3	Indirect through Number of O-Levels	Total Indirect Through Education
Effect on Time-Averaged Income	0.0033*	0.00003	0.00000	-0.00008	0.00014	0.00007
	(0.0018)	*(0.00008)*	*(0.0008)*	*(0.00018)*	(0.00022)	*(0.00034)*

p<0.1 = *, p<0.05 = **, p<0.01 = ***

Note: Bootstrapped (10,000 replications) standard errors in parentheses

For the male sample, there are lower coefficients for the (already non-significant) indirect pathways, particularly through education. There is a slightly stronger direct effect of receptive vocabulary on earnings than in the pooled sample, although the reduced sample size reduces the attributed level of significance to p=0.063, meaning we would be unable to reject the Null Hypothesis (β=0) associated with Research Question 1 at the 5% level in this sample.

Appendix Table A3 7 Results of model specification (removing female dummy variable from controls) ran only for female sample (n=1282)

	Direct Receptive Vocabulary	Indirect through Locus of Control	Indirect through CASMIN Level 2	Indirect through CASMIN Level 3	Indirect through Number of O-Levels	Total Indirect Through Education
Effect on Time-Averaged Income	0.0022	0.00003	0.00008	0.00057	0.00013	0.00081
	(0.0022)	*(0.00009)*	*(0.00018)*	*(0.00060)*	(0.0002)	*(0.00066)*

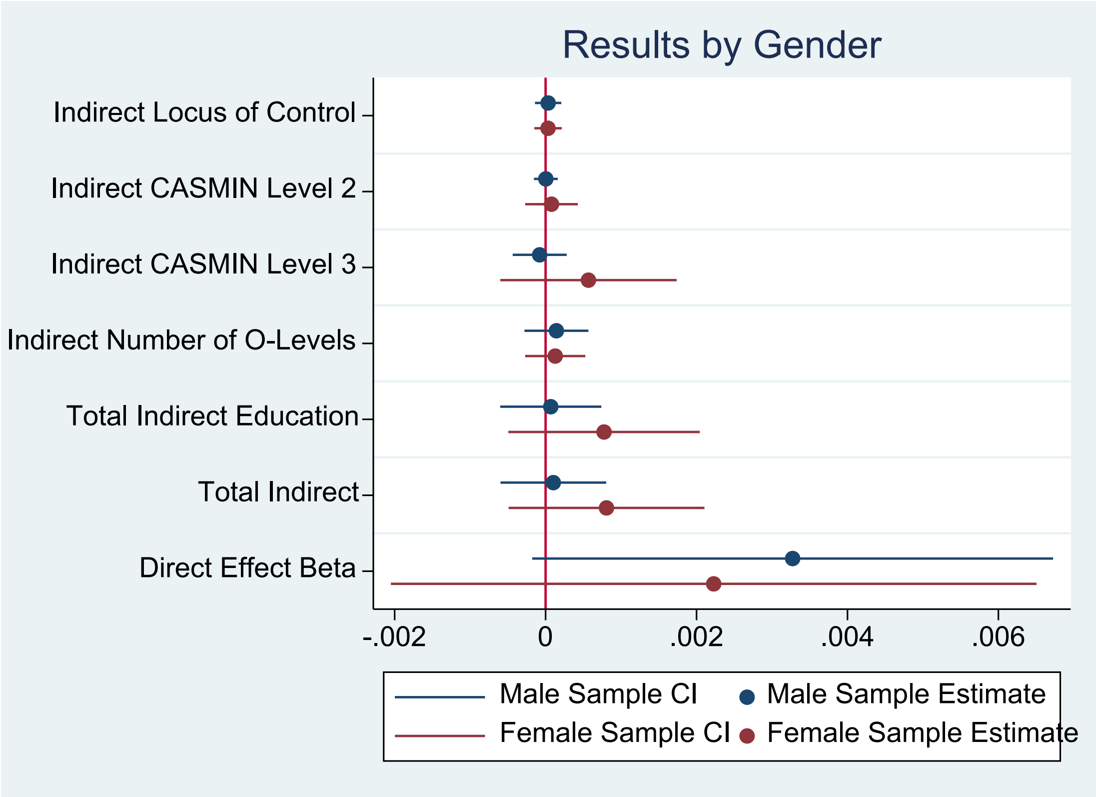
p<0.1 = *, p<0.05 = **, p<0.01 = ***

Note: Bootstrapped (10,000 replications) standard errors in parentheses

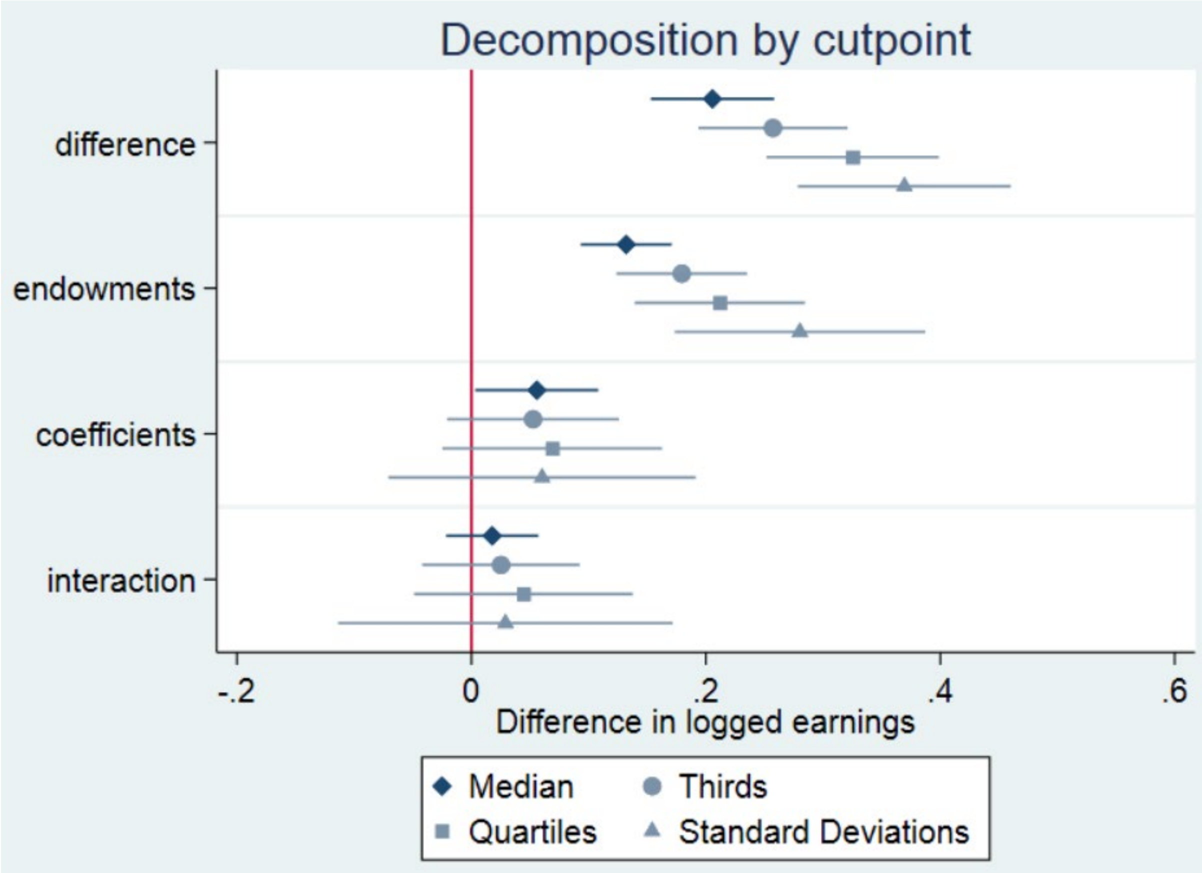
For the female sample, the indirect pathways are slightly stronger, particularly through education (p=0.231 for total education effect) although still insignificant. The estimated direct effect is weaker in the female sample by around a third, although the two β estimates are not significantly different from one another or from zero so this could just be noise in the data.

As Figure A3.7 shows, there are differential effects by gender but these are not significant enough to interpret in any meaningful way.

To draw conclusions about the effects of vocabulary on income, the pooled sample gives a better estimate of the varied effects but there is some justification for further research into differential effects and the potential for receptive vocabulary to be a stronger predictor of later income amongst the male sample. Whether this potentially stronger prediction is due to differences in the reliability of the initial estimate of receptive vocabulary (males score higher, suggesting they answer a greater number of questions on the test which improves our estimate of their underlying vocabulary) or reliability of the time-averaged estimate of logged income (males were more likely to be close to their maximum earning potential in this era due to a higher full-time employment rate) cannot be separated with the data available.

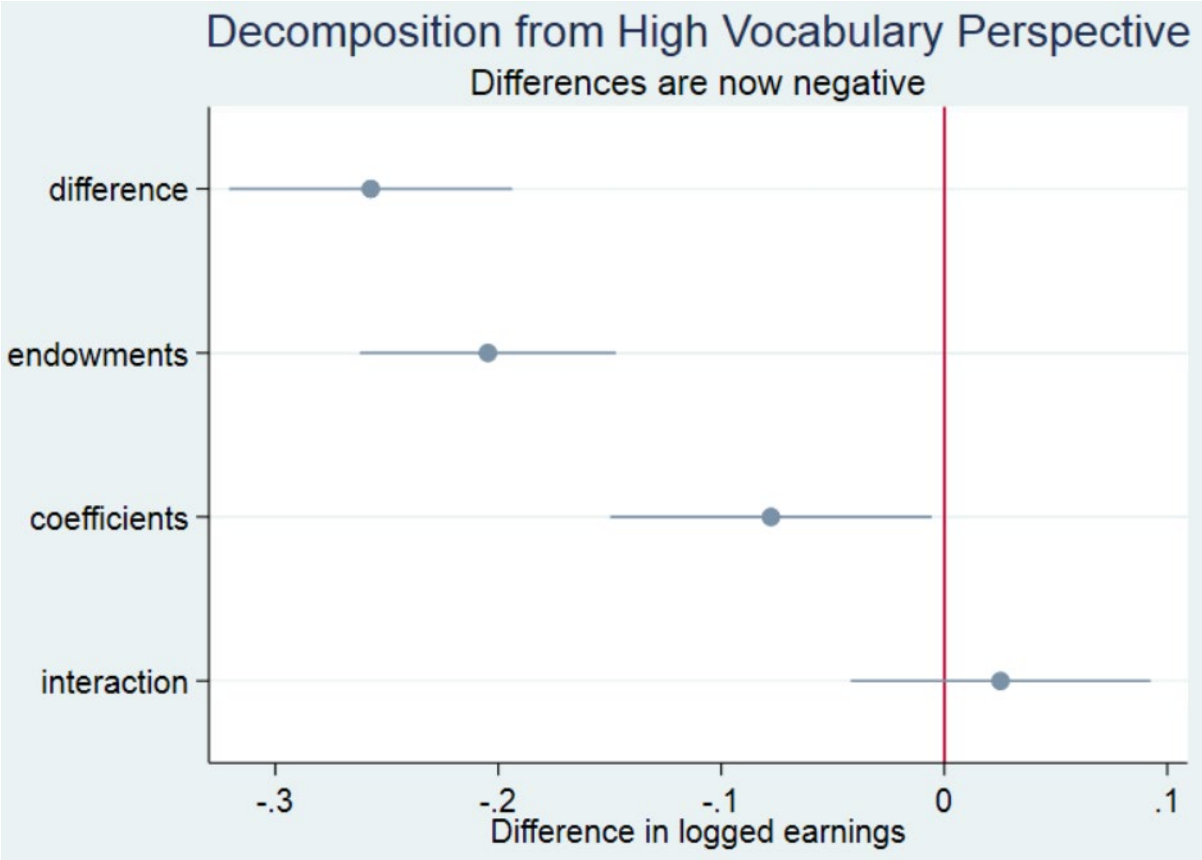


Appendix Figure A3 7- Indirect and Direct Effects, estimates and confidence intervals (CI), split by sample. As the confidence intervals show, there are no significant differences between males and females in their estimated effects.



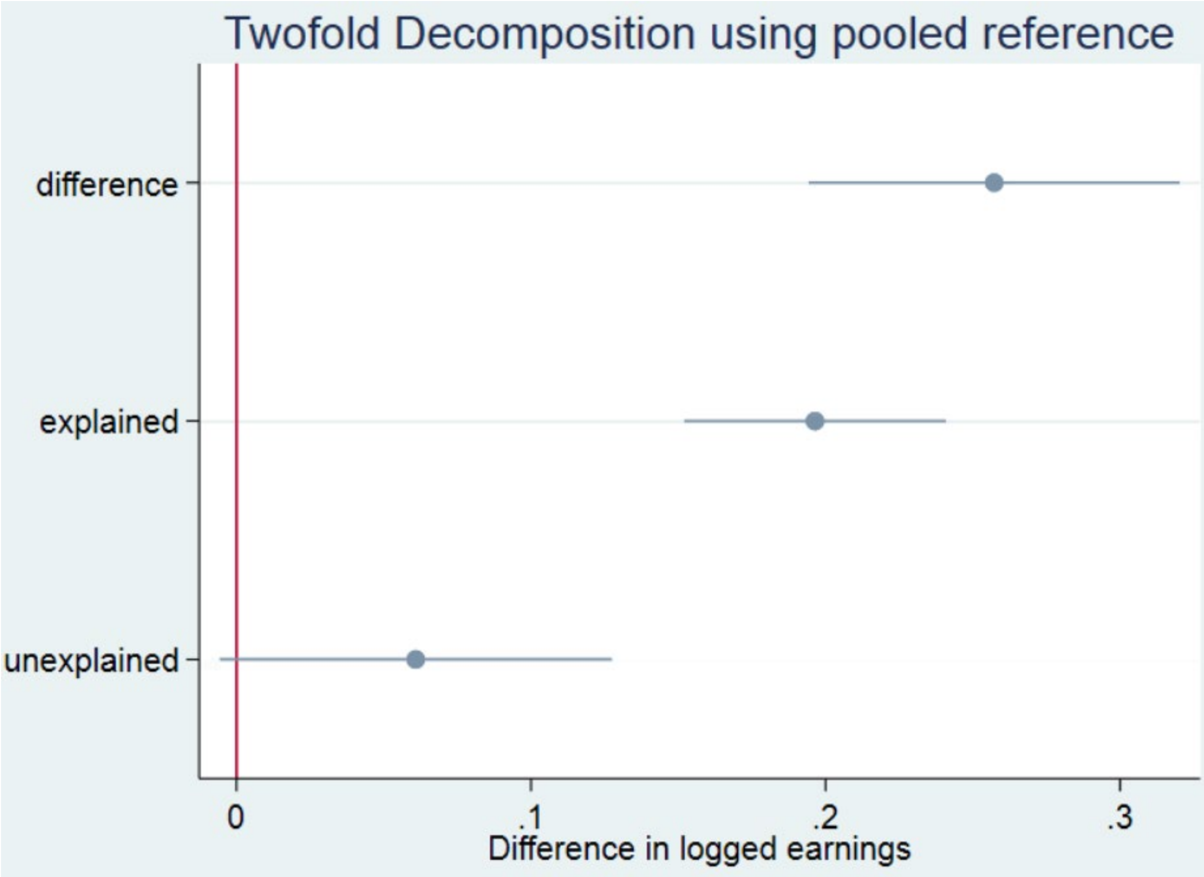
Appendix Figure A4 1- Decomposition by cut-point

As shown above, the decision of the cut-point has very little impact on our findings. When the two groups are distanced further apart in terms of the distribution of vocabulary scores we see a larger overall difference. This is largely picked up by the endowment effect. Therefore, our interpretation of the endowment effect is the most conservative available.



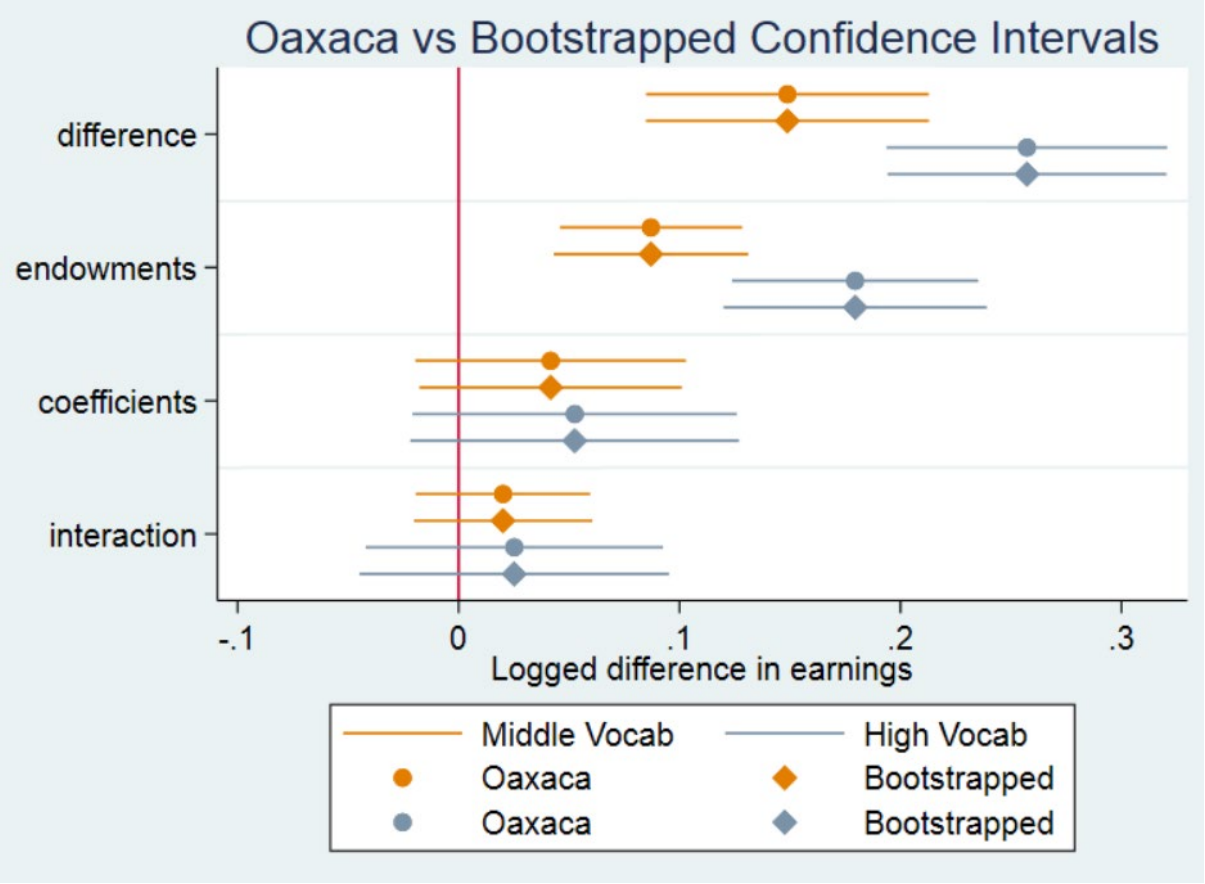
Appendix Figure A4 2- Decomposition from High Vocabulary Perspective

Reversing the perspective would make our coefficient difference significant. However, the interaction effect remains unchanged, even in direction, from our original measure. Therefore, the two “treatment” effects would be jointly insignificant and our summary that vocabulary does not significantly moderate the returns to other variables still holds.



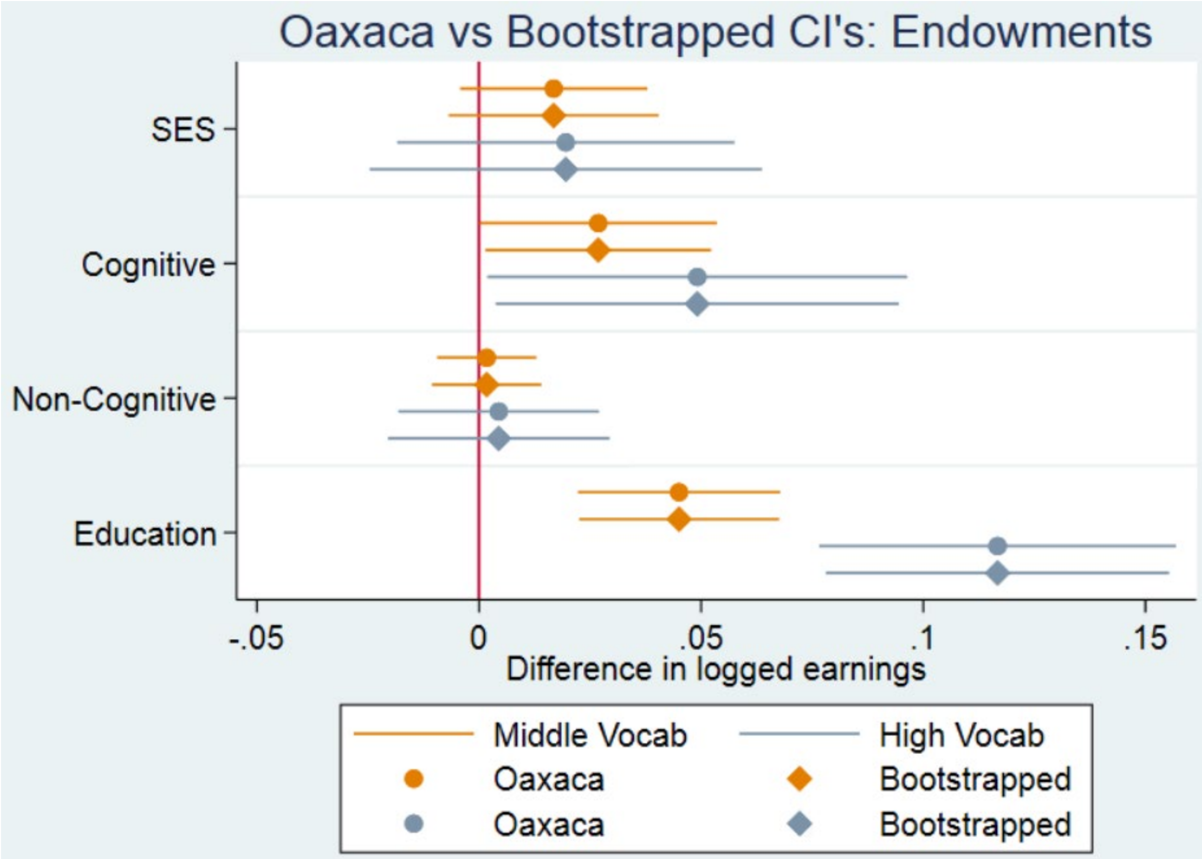
Appendix Figure A4 3- Twofold Decomposition using Pooled Reference

In this form of the decomposition, taken from Blinder (1973), the “treatment” effect is our unexplained component. This is still positive yet insignificant supporting our conclusion that vocabulary does not significantly moderate the returns to the other covariates.



Appendix Figure A4 4- Oaxaca vs Bootstrapped Confidence Intervals

Findings are robust to bootstrapped standard errors as opposed to the stata-produced Oaxaca standard errors.



Appendix Figure A4 5- Oaxaca vs Bootstrapped CI's: Endowments

Findings are robust to bootstrapped standard errors as opposed to the stata-produced Oaxaca standard errors.

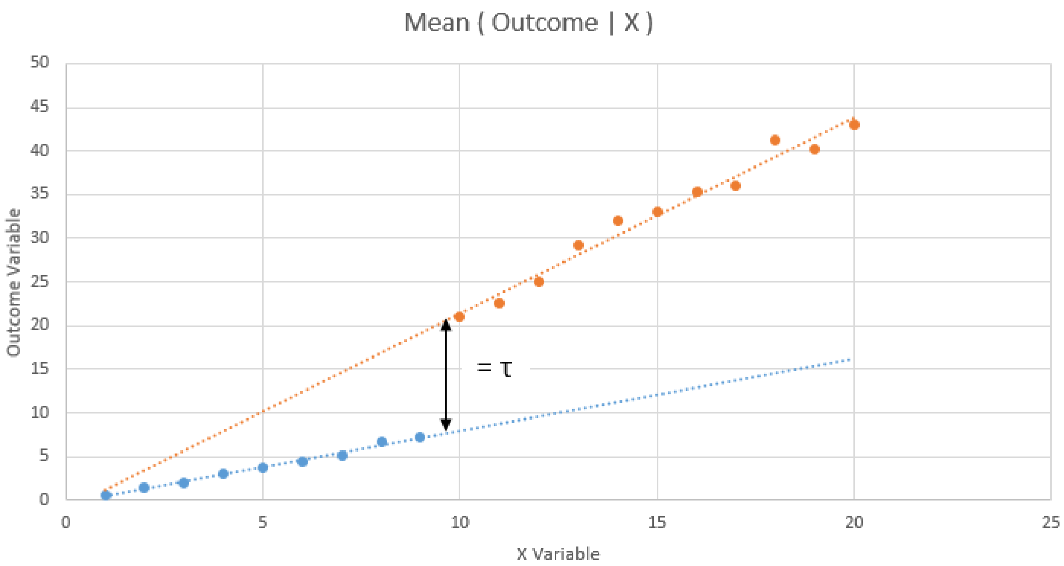
Chapter 5 Appendices

Appendix Table A5 1-

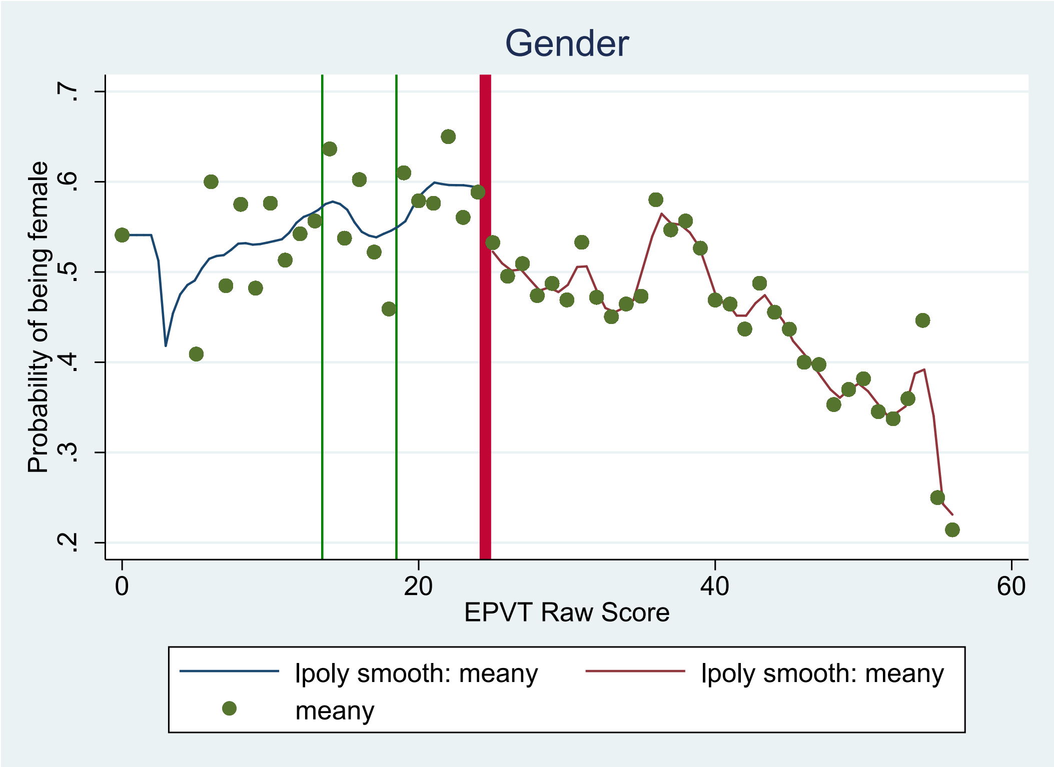
Background Variable	Differences found by bin width 2	3	4
Father SES	Cannot Reject H_0	Cannot Reject H_0	Cannot Reject H_0
Parental Social Class	Cannot Reject H_0	Cannot Reject H_0	Reject H_0
Gender	Cannot Reject H_0	Reject H_0	Reject H_0
Maternal Education	Cannot Reject H_0	Cannot Reject H_0	Cannot Reject H_0
Parental Education	Cannot Reject H_0	Cannot Reject H_0	Cannot Reject H_0
Singleton	Cannot Reject H_0	Cannot Reject H_0	Cannot Reject H_0
Region	Cannot Reject H_0	Cannot Reject H_0	Cannot Reject H_0
Seen Speech Therapist before 4y.o	Cannot Reject H_0	Cannot Reject H_0	Cannot Reject H_0
Gestation Period	Cannot Reject H_0	Cannot Reject H_0	Cannot Reject H_0

Null Hypothesis (H_0) – The two groups are statistically the same. $\frac{(\mu_{TI-LD} - \mu_{TI-NLD})}{s.e.} = 0$

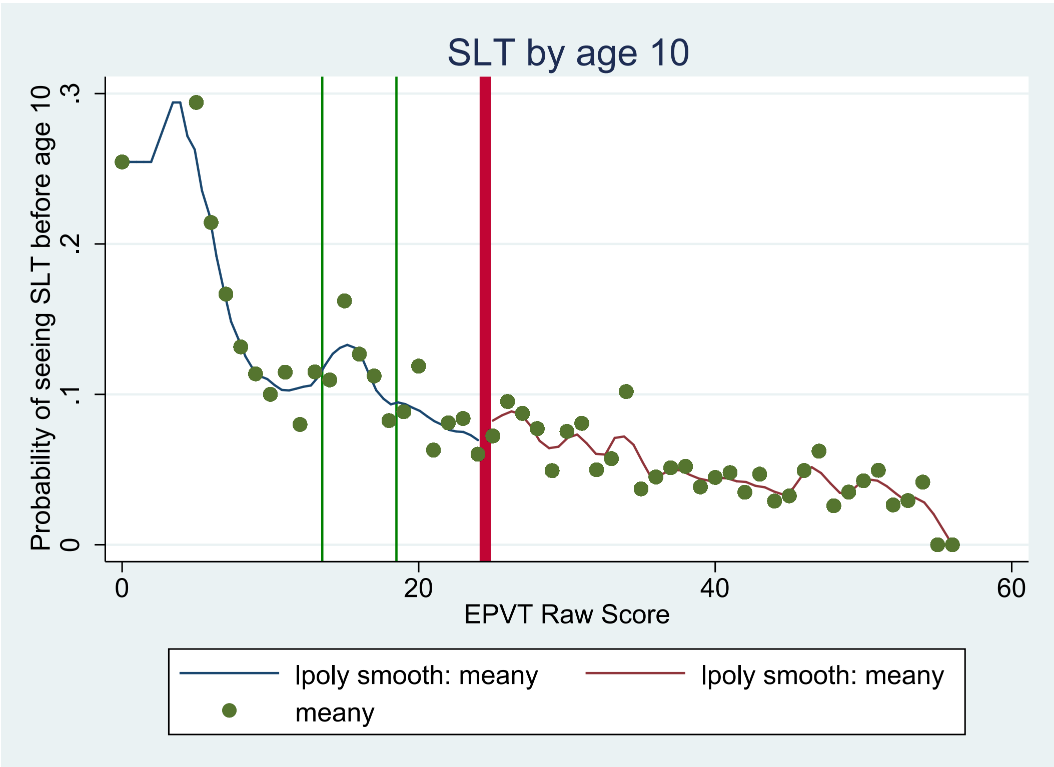
A



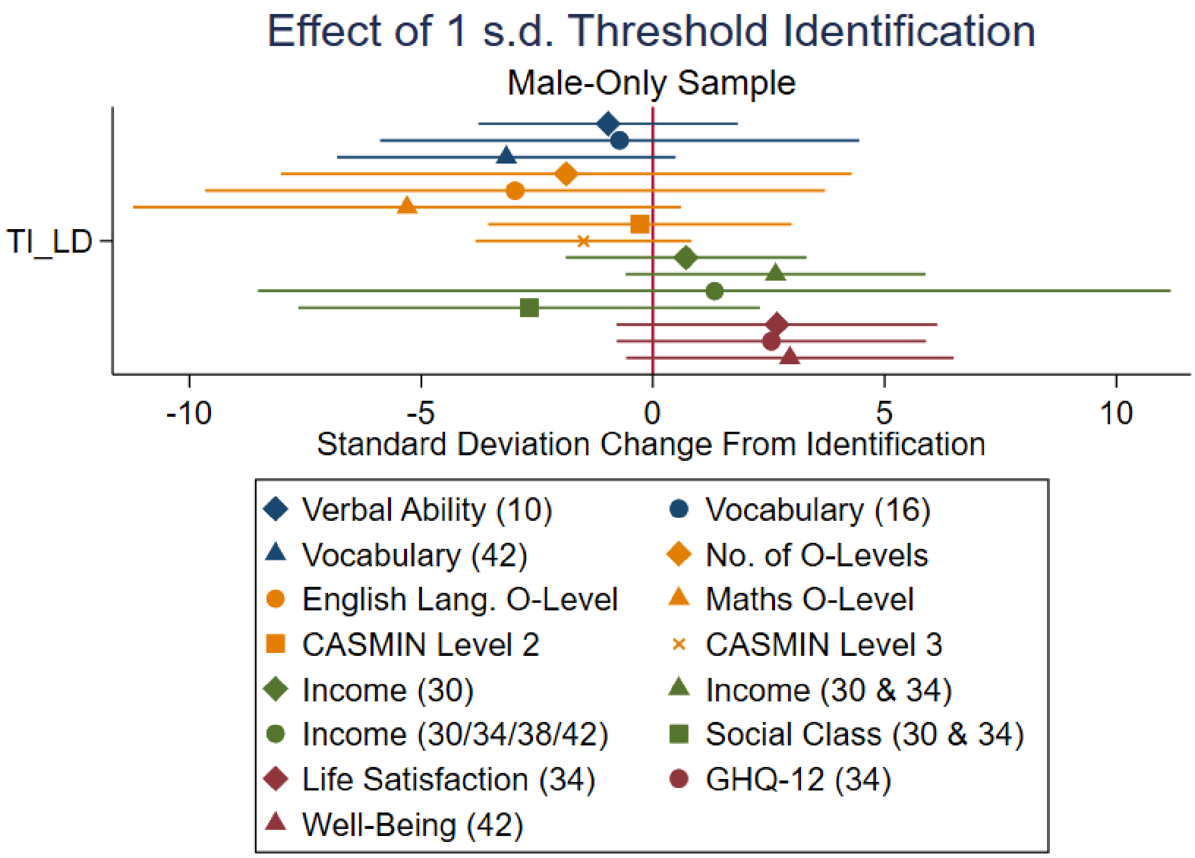
Appendix Figure A5- 1 Illustration of observable discontinuity effect at a threshold



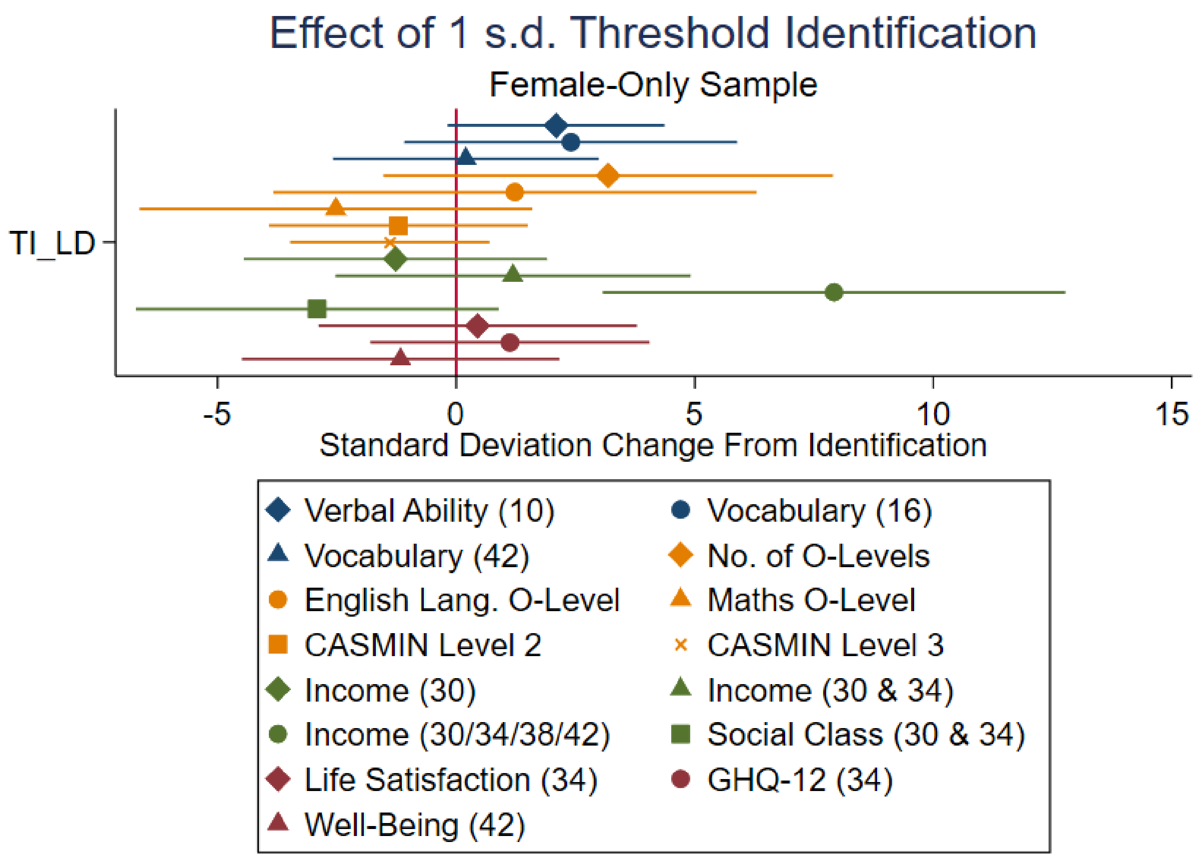
Appendix Figure A5- 2 Plot of proportion of females within each EPVT score bin



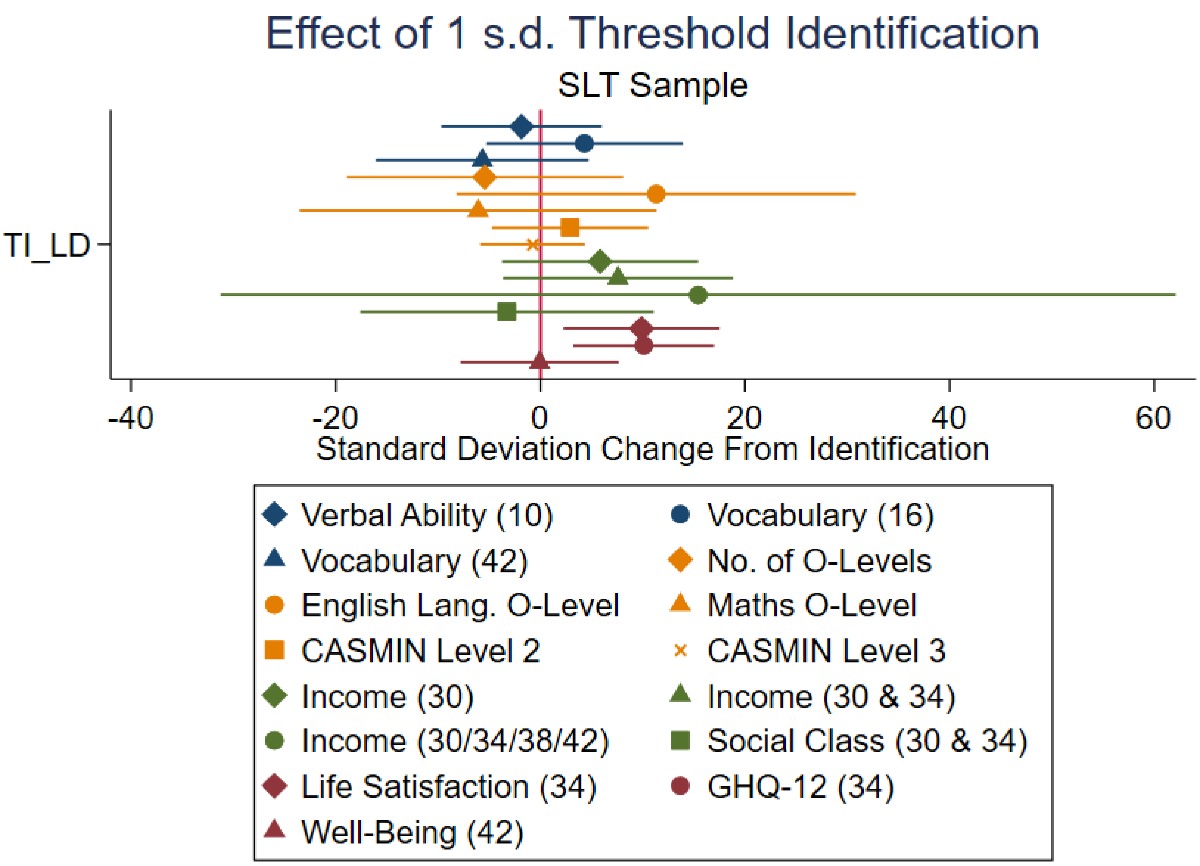
Appendix Figure A5- 3 Plot of SLT attendance by age 10 within each EPVT score bin



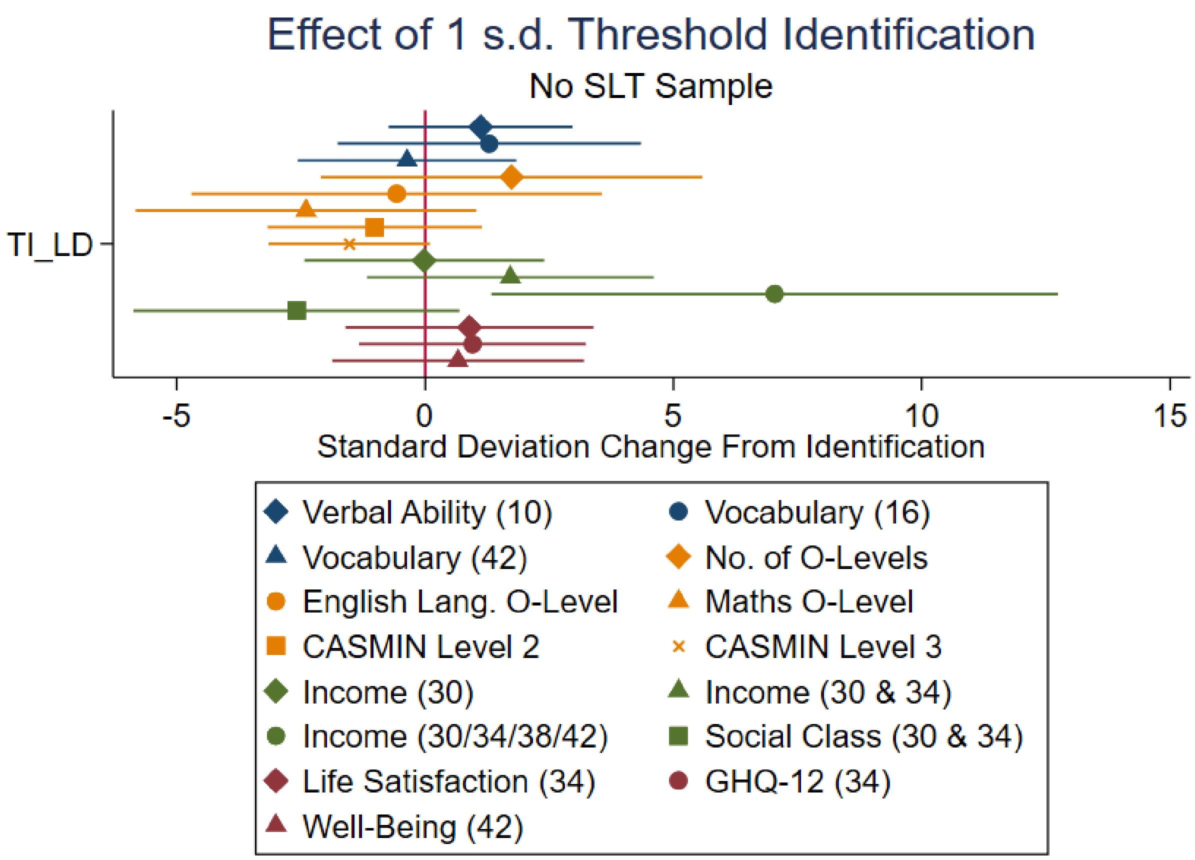
Appendix Figure A5 1 Stage 1 results for male-only sample



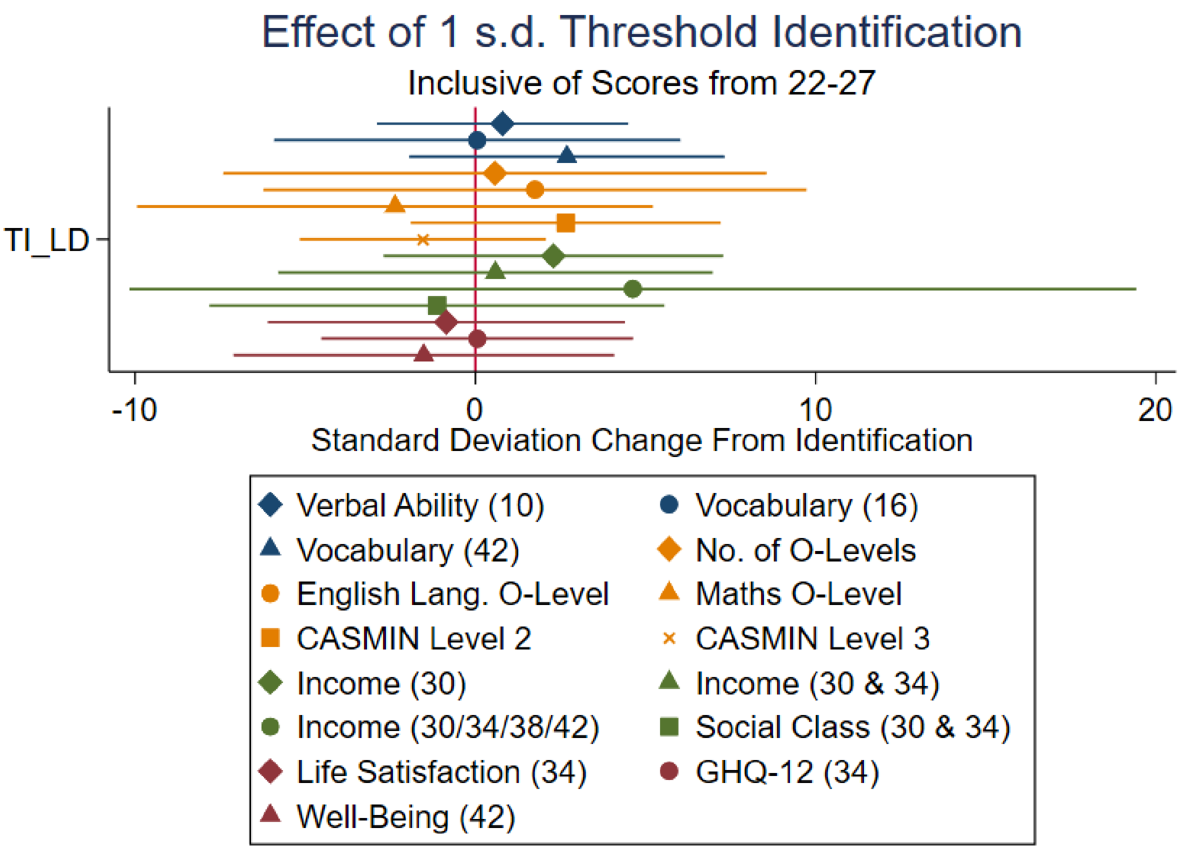
Appendix Figure A5 2-Stage 1 results for female-only sample



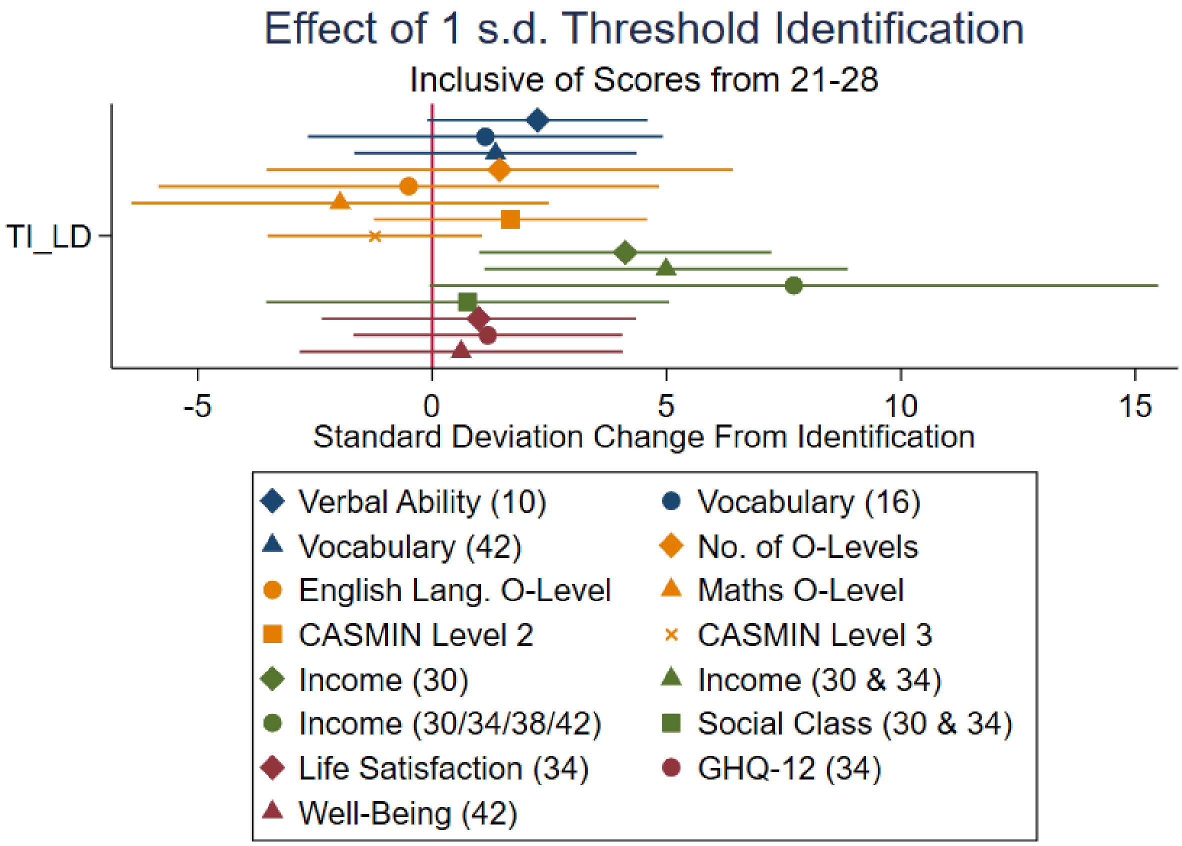
Appendix Figure A5 3- Stage 1 results for sample attending speech and language therapy



Appendix Figure A5 4- Stage 1 results for sample not attending speech and language therapy

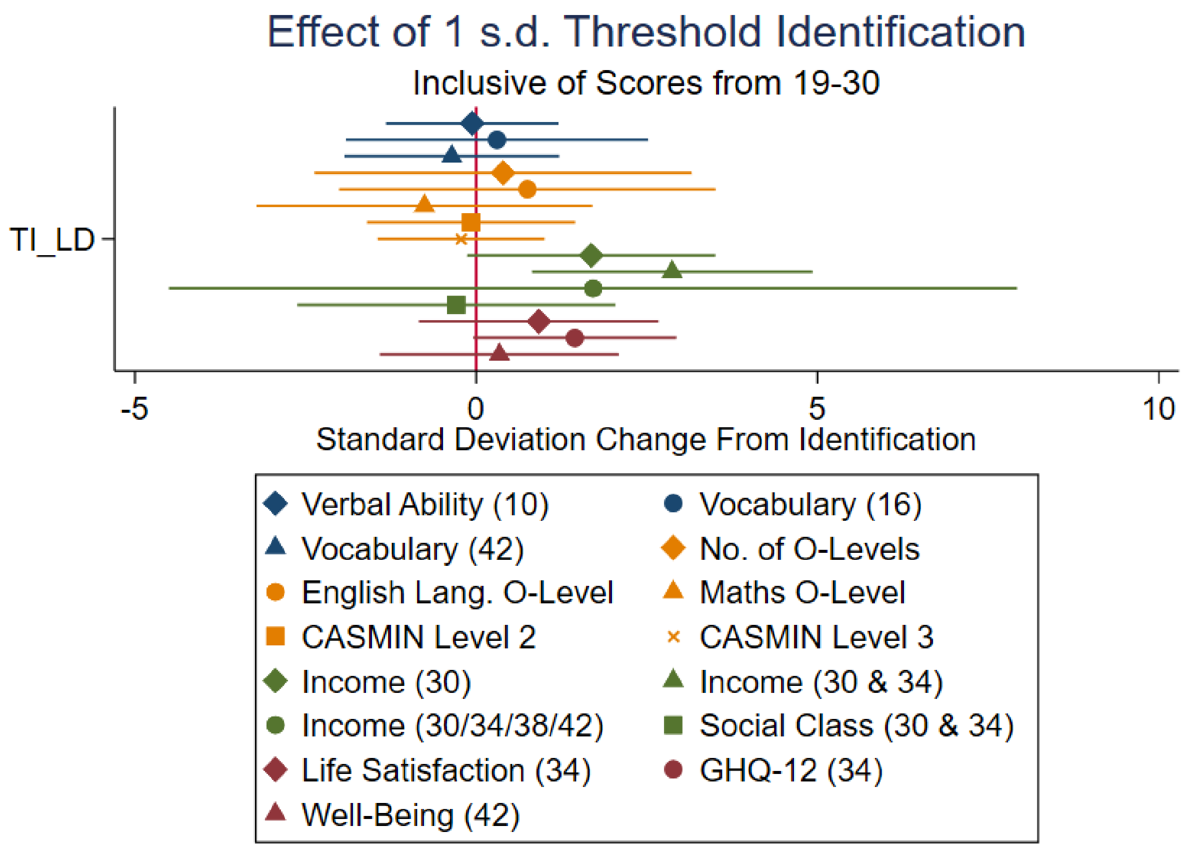


Appendix Figure A5 5 Stage 1 results for inclusion width 4 either side of threshold



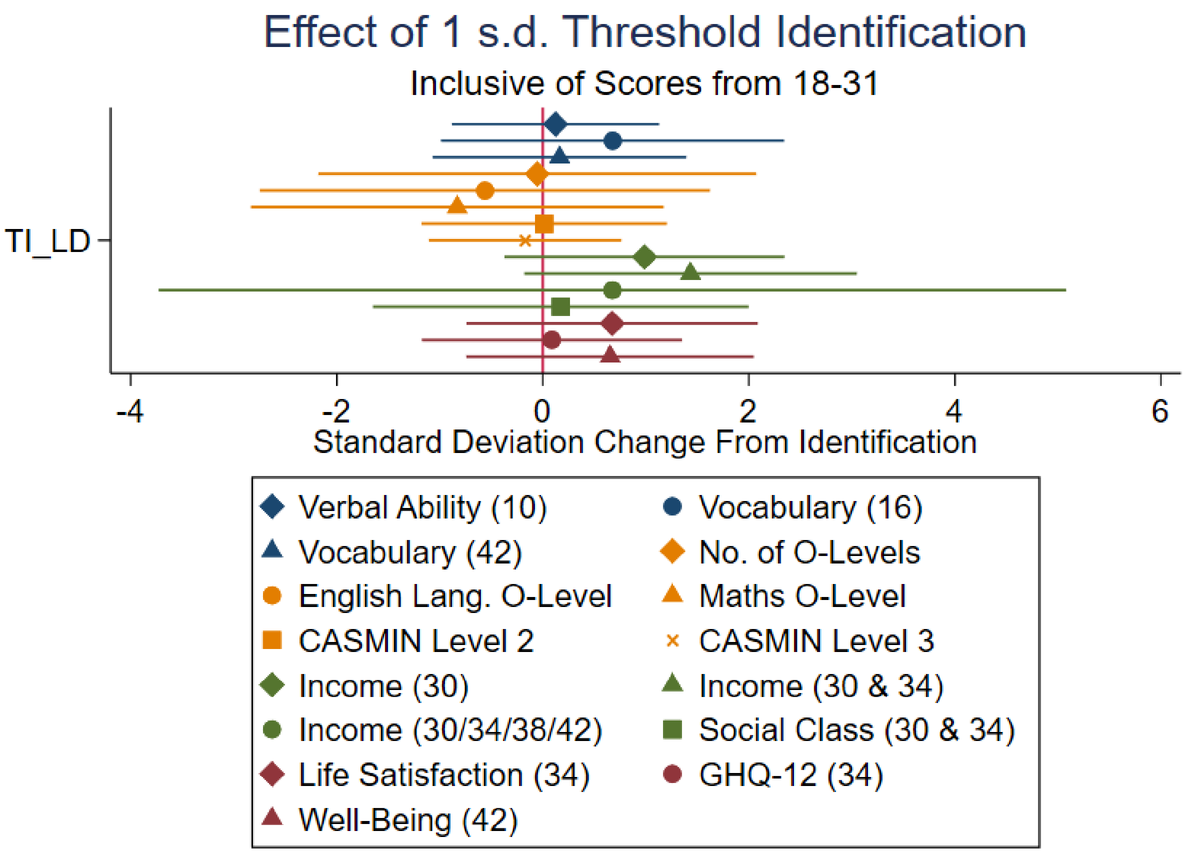
Appendix Figure A5 6 Stage 1 results for inclusion width 4 either side of threshold

Note: Positive scores also indicate that the threshold did not accurately identify a disorder.



Appendix Figure A5 7 Stage 1 results for inclusion width 6 either side of threshold

Note: Inclusion width of 5 not plotted here as this was used in the main text results.



Appendix Figure A5 8 Stage 1 results for inclusion width 7 either side of threshold

References

Aldridge S. (2003) The facts about social mobility. *New Economy* 10:189e93

Al Snih S., Markides K.S., Ray L., Ostir G.V. and Goodwin J.S. (2002). Handgrip strength and mortality in older Mexican Americans, *Journal American Geriatric Society*, vol. 50 (pg. 1250-56)

Almlund, M., Duckworth, A., Heckman, J. and Kautz, T. (2011) "Personality Psychology and Economics." *In Handbook of the Economics of Education,* edited by E. A. Hanushek, S. Machin, and L. W¨oßmann, Amsterdam: Elsevier, volume 4, 1–181.

Almond, D. Currie, J. and Duque, V. (2018) Childhood circumstances and adult outcomes: act ii*. J. Econ. Lit.,* 56 (4), pp. 1360-1446

American Speech-language-Hearing Association (ASHA) website: Undated publication. https://www.asha.org/public/speech/development/language_speech/ *Last accessed 22/08/2020*

Angrist, J. and Krueger, A. (2001) "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," Journal of Economic Perspectives, Fall 2001, 15 (4), 69–85.

Angrist, J. D. and Pischke, J. (2009). Mostly harmless econometrics: An Empiricist's Companion. Princenton: Princenton University Press.

Angrist, J. D. and Pischke, J. (2015). Mastering metrics: The path from cause to effect. Princeton University Press.

Archer, W., and J. Davison. (2008). Graduate employability: What do employers think and want? The Council for Industry and Higher Education. http://www.cihe-uk.com (accessed June 18, 2011).

Ashenfelter, O., and Oaxaca, R. (1987) ''The Economics of Discrimination: Economists Enter the Courtroom,'' *American Economic Review*, May 1987, 77, 321–25.

Barnett S.W., Jung K., Yarosz D.J., Thomas J., Hornbeck A., Stechuk R., and Burns S. "Educational Effects of the Tools of the Mind Curriculum: A Randomized Trial." *Early Childhood Research Quarterly* 23, 3: (2008) 299–313.

Baron, R. and Kenny, D. (1986) The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, Vol 51(6), Dec 1986, 1173-1182

Barre N, Morgan A, Doyle LW, Anderson PJ. (2011) Language abilities in children who were very preterm and/or very low birth weight: a meta-analysis. *J Pediatr*.;158(5):766–774.e1. doi: 10.1016/j.jpeds.2010.10.032.

Barrick, M. R. and M. K. Mount (1991). The Big Five personality dimensions and job

performance: A meta-analysis. *Personnel Psychology* 44 (1), 1-26.

Basit, T.N., Hughes, A., Iqbal, Z., & Cooper, J. (2015). The influence of socio-economic status and ethnicity on speech and language development. *International Journal Of Early Years Education*, 23(1), 115-133. doi: 10.1080/09669760.2014.973838

Beard, A. (2018). Speech, language and communication: a public health issue across the lifecourse. *Paediatr. Child Health*. 28, 126–131. doi: 10.1016/j.paed.2017.12.004

Becker, G. S. (1964): Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education. *Columbia University Press*, New York.

Becker, G. (1968), "Crime and Punishment: An Economic Approach", *The Journal of Political Economy*, 76, 169–217.

Beitchman J.H, Brownlie, E.B., Walters H., Inglis A. and Lancee W. (1996). Long-Term Consistency in Speech/Language Profiles: II. Behavioral, Emotional, and Social Outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(6), 815–825. https://doi.org/https://doi.org/10.1097/00004583-199606000-00022

Beitchman JH, Wilson B, Johnson CJ, Atkinson L, Young A, Adlaf A, Escobar M, Douglas L. (2001) Fourteen-year follow-up of speech/language-impaired and control children: Psychiatric outcome. *Journal of the American Academy of Child and Adolescent Psychiatry* ;40(1):75-82.

Beitchman, J., & Brownlie, E. (2010). Language development and its impact on children's psychological and emotional development. In R. Tremblay, R. Barr, & R. Peters (Eds.). *Encyclopedia on early childhood development* (online). Montreal, Quebec: Centre of Excellence for Early Childhood Development. Available online at: http://www.childencyclopedia.com/Pages/PDF/Beitchman-BrownlieANGxp_rev.pdf

Beitchman J, Brownlie EB. (2013) Language Disorders in Children. *Vol. 28. Oxford*: Hogrefe

Belbasis L, Savvidou MD, Kanu C, Evangelou E and Tzoulaki I (2016) Birth weight in relation to health and disease in later life: an umbrella review of systematic reviews and meta-analyses. *BMC Med* ;14:147.Belfer, M (2008). Child and adolescent mental disorders: The magnitude of the problem across the globe. *Journal of Child Psychology and Psychiatry*, 49 (2008), pp. 226–236

Benner G.J. (2002) Language skills of children with EBD: A Literature Review. *Journal of Emotional and Behavioural Disorders.* 10(1)43–56.

Bercow, J. (2008). The Bercow Report: A review of services for children and young people (0-19) with speech, language and communication needs*. Nottingham: DCSF.*

https://www.education.gov.uk/publications/eOrderingDownload/Bercow-Report.pdf

Bercow, J (2018) Bercow: Ten years on. An independent review of provision for children and young people with speech, language and communication needs in England *Retrieved from www.bercow10yearson.com*

Bishop DVM (2002) The role of genes in the etiology of specific language impairment. *Journal of Communication Disorders.* 35:311-328

Bishop, DVM, Snowling, M., Thompson, P. A., Greenhalgh, T. and CATALISE consortium (2016) 'CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children', *PLOS one.*

Bishop DVM, Snowling MJ, Thompson PA, Greenhalgh T. (2017) Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: terminology. *Journal of Child Psychology and Psychiatry.* 58: 1068–80

Bjorklund A, Jantti M, Solon G. (2007) Nature and Nurture in the Integenerational Transmission of Socioeconomic Status: Evidence from Swedish Children and Their Biological and Rearing Parents. *BE Journal of Economic Analysis & Policy*. 7: Iss, 2 (Advances), Article 4

Blanden, J. (2013) 'Cross-National Rankings of Intergenerational Mobility: A Comparison of Approaches from Economics and Sociology', *Journal of Economic Surveys*, 27, 38-74.

Blanden, J. (2015) 'Intergenerational Income Persistence', *Data and Measurement Series, IZA World of Labor*.

Blanden, J. and Gibbons, S. (2006) The persistence of poverty across generations: A view from two British cohorts, *Bristol: The Policy Press*.

Blanden, J., Goodman A., Gregg P. and Machin S. (2004) 'Changes in Intergenerational Mobility in Britain', *in M. Corak (ed.) Generational Income Mobility*, Cambridge University Press.

Blanden, J; Gregg, P. and Macmillan, L. (2006) "Explaining Intergenerational Income Persistence: Non-Cognitive Skills, Ability, and Education," *CMPO Working Paper #06/146, Dept. of Economics*, University of Bristol.

Blanden, J., & Machin, S. (2010). Intergenerational inequality in Early Years assessments. In K. Hansen, H. Joshi, & S. Dex (Eds.), *Children of the 21st Century: The First Five Years* (pp. 153–168). *Policy Press*

Blanden, J., Gregg, P and Macmillan, L. (2013) 'Intergenerational Persistence in Income and Social Class: The Impact of Within-Group Inequality', *Journal of Royal Statistical Society: Series A*, 176, 541-63

Bleakley, H; Chin, A (2002). Language Skills and Earnings: Evidence from Childhood Immigrants*. http://www.mitpressjournals.org/doi/pdf/10.1162/003465304323031067*

Bolton, P. (2012): Education: historical statistics*. House of Commons Library*

Borghans, L., Meijers, H., and ter Weel, B. (2008) "The Role of Noncognitive Skills in Explaining Cognitive Test Scores." *Economic Inquiry* 46, 1: 2–12.

Bornstein, M. H., Hahn, C. S. and Putnick, D. L. (2016) ‘Long-term stability of core language skill in children with contrasting language skills’, *Developmental Psychology*, 52 (5), pp. 704–16.

Bornstein, M. H., Hahn, C.-S., Putnick, D. L., & Pearson, R. M. (2018). Stability of core language skill from infancy to adolescence in typical and atypical development. *Science Advances*, 4(11), 31 eaat7422. https://doi.org/10.1126/sciadv.aat7422

Botting, N., & Conti-Ramsden, G. (2000). Social and behavioural difficulties in children with language impairment. *Child Language Teaching and Therapy*. https://doi.org/10.1177/026565900001600201

Bound, J, Brown C, and Mathiowetz N. (2001) “Measurement Error in Survey Data.” In *Handbook of Econometrics*, edited by James J. Heckman, and Edward Leamer, Amsterdam: Elsevier Science, volume 5 of *Handbooks in Economics*, 3705–3843

Bowen, C. (2009) Children's Speech Sound Disorders *(Chichester: Wiley-Blackwell).*

Boyle J, McCartney E, Forbes J, O'Hare A (2007) A randomised controlled trial and economic evaluation of direct versus indirect and individual versus group modes of speech and language therapy for children with primary language impairment. *Health Technol Assess* 11: iii–iv, xi.

Bradley, R. (2002). Environment and Parenting. In M. Bornstein (Ed.) *Handbook of parenting, 2nd Ed.* Hillsdale, N. J.: Lawrence Erlbaum Associates.

Brauns H., Scherer S., Steinmann S. (2003) The CASMIN Educational Classification in International Comparative Research. In: Hoffmeyer-Zlotnik J.H.P., Wolf C. (eds) *Advances in Cross-National Comparison. Springer*, Boston, MA

Brimer and Dunn (1962): Manual for the English Picture Vocabulary Tests; Distributed by *Educational Evaluation Enterprises*, Bristol.

Brinton, B and Fujiki, M (2010). Living with language impairment. *Journal of Interactional Research in Communication Disorders,* 1, pp. 69–94

Brinton, B. Fujiki, M. Montague, E.C., Hanton, J.L. (2000). *Children with language impairment in cooperative work groups: A pilot study.* Language, Speech, and Hearing Services in Schools, 31, pp. 252–264

Broomfield, J., and Dodd, B. (2004). Children with speech and language disability: caseload characteristics. *International Journal of Language and Communication Disorders*, 39 (3), pp. 303-24. Cited in Law et al. (2017) Language as a child wellbeing indicator. Available at: https://www.eif.org.uk/report/language-as-a-child-wellbeing-indicator

Brownlie, E.B., Beitchman, J.H., Escobar, M., Young, A., Atkinson, L., Johnson, C., Wilson, B., Douglas, L. (2004) Early Language Impairment and Young Adult Delinquent and Aggressive Behavior. *J Abnorm Child Psychol* 32, 453–467 (2004). https://doi.org/10.1023/B:JACP.0000030297.91759.74

Brunello G, Fort M, Schneeweis N, Winter-Ebmer R. (2015) The causal effect of education on health: what is the role of health behaviors? *Health Economics*.

Boyd, A., Golding, J., Macleod, J., Lawlor, D., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S. and Smith, G.D. (2013) Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children*, International Journal of Epidemiology,* Volume 42, Issue 1, February 2013, Pages 111–127, https://doi.org/10.1093/ije/dys064

Byers, R., McLaughlin, C. and Peppin-Vaughan, R., (2012), The research perspective: vulnerability and prevalence. In C. McLaughlin, R. Byers and C. Oliver (eds), *Perspectives on Bullying and Difference: Supporting Young People with Special Educational Needs and/or Disabilities in Schools (London: National Children's Bureau).*

Camacho C, Straatmann VS, Day JC, Taylor-Robinson D (2019) Development of a predictive risk model for school readiness at age 3 years using the UK Millennium Cohort Study. *BMJ Open*;9:e024851. doi:10.1136/bmjopen-2018-024851

Carnevale, A.P., Smith, N., and Strohl, J. (2013). Recovery: Job Growth and Education Requirements Through 2020. *Georgetown University Center on Education and the Workforce*.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., GoldinMeadow, 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 of the United States of America,* 110, 11278– 11283. https://doi.org/10.1073/pnas.1309518110

Cartmill, E.A. (2016). "Mind the Gap: Assessing and Addressing the Word Gap in Early Education." *Policy Insights from the Behavioral and Brain Sciences* 3 (2): 185–193.

Case, A., Lubotsky, D., and Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *American Economic Review*, 92(5):1308–1334.

Cherlin, AJ; Furstenberg, FF Jr; Chase-Lansdale, L; Kiernan, KE; Robins, PK; Morrison, DR; Teitler, JO (1991). Longitudinal studies of effects of divorce on children in great britain and the United States. *Science,* 252(5011), 1386. Retrieved from http://search.proquest.com/docview/213571295?accountid=12753

Choudhury N, Benasich AA (2003): A family aggregation study: the influence of family history and other risk factors on language development. *Journal of Speech Language and Hearing Research*; 46:261-272

Chowdry, H., Crawford, C., Dearden, L., Goodman, A. and Vignoles, A. (2008) Widening Participation in Higher Education: Analysis using Linked Administrative Data, *London: Institute for Fiscal Studies.*

Christensen, D., Zubrick, S. R., Lawrence, D., Mitrou, F. and Taylor, C. L. (2014) 'Risk factors for low receptive vocabulary abilities in the preschool and early school years in the longitudinal study of Australian children', *PLoS ONE*, 9 (7).

Christensen D, Taylor CL, Zubrick SR (2017) Patterns of Multiple Risk Exposures for Low Receptive Vocabulary Growth 4–8 Years in the Longitudinal Study of Australian Children. *PLoS ONE* 12(1): e0168804. pmid:28114381

Clark, D. and Royer, H. (2013) "The Effect of Education on Adult Mortality and Health: Evidence from Britain," *American Economic Review*, October 2013, 103 (6), 2087–2120.

Clegg J, Hollis C and Rutter M (1999) 'Life Sentence: what happens to children with developmental language disorders in later life?' in *RCSLT Bulletin. Royal College of Speech and Language Therapists.*

Clegg, J., Law, J., Rush, R., Peters, T.J. and Roulstone, S. (2015), The contribution of early language development to children's emotional and behavioural functioning at 6 years: an analysis of data from the Children in Focus sample from the ALSPAC birth cohort. *Journal of Child Psychology and Psychiatry*, 56: 67-75. doi:10.1111/jcpp.12281

Clotfelter, C.T., Ladd, H.F., Vigdor, J.L., (2009). The academic achievement gap in grades 3 to 8. *The Review of Economics and Statistics* 91, 398-419

CLS (Centre for Longitudinal Studies) Caraloc Questionnaire for Locus of Control: https://cls.ucl.ac.uk/cls-studies/1970-british-cohort-study/bcs70-age-10-sweep/ (retrieved 15/03/2020)

Cobb-Clark, D; Schurer, S (2012). The stability of big-five personality traits. *Economics Letters*. Volume 115 (1) p11-15

Cobb-Clark, D.A., Salamanca, N., Zhu, A., (2019). Parenting Style as an Investment in Human Development. *Journal of Population Economics* 32, 1315–1352

Cocking, R., & Mestre, J. (1988). Introduction: Considerations of language mediators of mathematics learning. In R. Cocking & J. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 3–16). Hillsdale, NJ: Erlbaum.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: *Lawrence Erlbaum Associates*, Publishers.

Cole, K (2006). Wellbeing, psychological capital and unemployment: An Integrated Theory. *https://melbourneinstitute.com/downloads/hilda/Bibliography/Conference_Papers/Paris%20Paper%2020061.pdf*

Conger, R.D., Ge, X., Elder, G.H., Lorenz, F.O and Simons, R.L. (1994), Economic stress, coercive family process, and developmental problems of adolescents. *Child Development*, 65, 541–61.

Conti, G., Heckman, J. J., and Urzua, S. (2010). The education-health gradient. *American Economic Review: Papers & Proceedings,* 100(2):234–238.

Conti, G., Heckman, J. J., and Urzua, S. (2011). Early endowments, education, and health. Working Paper 2011-001, *Human Capital and Economic Opportunity: A Global Working Group*

Conti, G. Mason, G., Poupakis, S. (2019) Developmental origins of health inequality. *Oxford Research Encyclopedia of Economics and Finance*

Conti-Ramsden, G., & Botting, N. (2004). Social difficulties and victimization in children with SLI at 11 years of age. *Journal of Speech, Language, and Hearing Research*, 47, 145–161

Conti-Ramsden, G., Mok, P., Pickles, A., Durkin, K. (2013) Adolescents with a history of specific language impairment (SLI): Strengths and difficulties in social, emotional and behavioral functioning. *Research in Developmental Disabilities*, 34, pp. 4161-4169, 10.1016/j.ridd.2013.08.043

Conti-Ramsden G, Durkin K (2016) What Factors Influence Language Impairment Considering Resilience as well as Risk. *Folia Phoniatr Logop* 2015;67:293-299. doi: 10.1159/000444750

Conti-Ramsden, G., Durkin, K., Toseeb, U., Botting, N., & Pickles, A. (2018). Education and employment outcomes of young adults with a history of developmental language disorder. *International Journal of Language & Communication Disorders*, 53(2), 237–255. doi:10.1111/1460-6984.12338

L.J. Conway, P.A. Levickis, J. Smith, F. Mensah, M. Wake, S. Reilly (2018) Maternal communicative behaviours and interaction quality as predictors of language development: Findings from a community-based study of slow-to-talk toddlers. *International Journal of Language & Communication Disorders*, 53 (2) (2018), pp. 339-354, 10.1111/1460-6984.12352

Costa, P.T., and McCrae, R.R. (1992). "Four Ways Five Factors are Basic." *Personality and Individual Difference* 13, 6: 653–665. . Revised NEO Personality Inventory (NEO PI-R) and the NEO Five-Factor Inventory (NEO-FFI) professional manual. *Odessa, FL: Psychological Assessment Resources*.

Costa Dias, M., Parodi F., and Joyce R. (2018). "The Gender Pay Gap in the UK: Children and Experience in Work." *IFS Working Paper 18/02.* http://hdl.handle.net/10419/200291

Crawford, C. and Greaves, E.. (2015). Socio-Economic, Ethnic and Gender Differences in HE Participation. London: *Institute for Fiscal Studies*

Crawford C, Gregg P, Macmillan L, Vignoles A, Wyness G (2016) Higher education, career opportunities, and intergenerational inequality, *Oxford Review of Economic Policy*, Volume 32, Issue 4, WINTER 2016, Pages 553–575, https://doi.org/10.1093/oxrep/grw030

Crystal, D. (1997). The Cambridge encyclopedia of language, second edition. *Cambridge University Press*

Cunha, F, Heckman JJ, Lochner LJ, and Masterov DV (2006). “Interpreting the Evidence on Life Cycle Skill Formation.” In *Handbook of the Economics of Education*, edited by Eric A. Hanushek, and Frank Welch, Amsterdam: North-Holland, chapter 12, 697–812.

Cunha, F., Heckman, J.J., (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources* 43, 738-782

Cunha, F., Heckman, J.J., Schennach, S.M., (2010). Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica* 78, 883-931

Cunningham, A & Stavonich, K (1997) Early Reading Acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology* Vol. 33 (6) p934-945

Currie, J. and Almond, D. (2011). Human capital development before age five. In Card, D. and Ashenfelter, O., editors, *Handbook of Labor Economics,* volume 4, pages 1315–1486. Elsevier

Dale PS, Harlaar N, Hayiou-Thomas ME, Plomin R. (2010) The etiology of diverse receptive language skills at 12 years. *Journal of Speech, Language, and Hearing Research.* ;53:982–992

Dasu T, Johnson T (2003). Exploratory Data Mining and Data Cleaning. *Wiley-IEEE*.

Department for Education (2016), Special Educational Needs and Disability (SEND) (available at: https://www.gov.uk/government/policies/special-educational-needs-and-disability-send) Last accessed: 27/08/2019

Department for Education (2017). Unlocking Talent, Fulfilling Potential: A plan for improving social mobility through education. *ISBN 978-1-5286-0046-0* CCS0917075678 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/667690/Social_Mobility_Action_Plan_-_for_printing.pdf

Diamond A., Barnett S., Thomas J., and Munro S. (2007) “Preschool Program Improves Cognitive Control.” *Science 318*, 5855: 1387–1388.

Dionne G, Dale PS, Boivin M, Plomin R (2003) Genetic Evidence for Bidirectional Effects of Early Lexi- cal and Grammatical Development. *Child Development* 74: 394–412. PMID

Dockrell, J; Ricketts, J; Lindsay, G (2012). Understanding speech language and communication needs: Profiles of needs and provision.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/219630/DFE-RR247-BCRP4.pdf

Dockrell, J., Lindsay, G., Roulstone, S. and Law, J., (2014). Supporting children with speech, language and communication needs: an overview of the results of the Better Communication Research Programme. *International Journal of Language & Communication Disorders*, *49*(5), pp.543-557. DOI/10.1111/1460-6984.12089

Dockrell, J, Lindsay, G, Palikara, O and Cullen, M.A. (2007). Raising the achievements of children and young people with specific speech and language difficulties and other special educational needs through school to work and college *(London: Department for Education and Skills).*

Dockrell, J. E., and Marshall, C. (2015). Measurement Issues: Assessing language skills in young children. *Child and Adolescent Mental Health*, 20, pp. 116-25.

Dodd, B. (2005) Differential Diagnosis and Treatment of Children with Speech Disorder, 2nd edition *(London: Whurr).*

Dolton, P., & Vignoles, A. (2000). The incidence and effects of overeducation in the UK graduate labor market. *Economics of Education Review*, 19(2), 179–198.

Drillien CM. (1967) The incidence of mental and physical handicaps in school age children of very low birth weight. II. *Pediatrics.*, 39: p238–47.

Dunn, D. & Dunn, L (2009); Manual for the British Picture Vocabulary Scales – III. *National Foundation for Educational Research in England and Wales*, GL Assessment (Firm)

Durham, R. E., Farkas, G., Hammer, C. S., Tomblin, J. B., & Catts, H. W. (2007). Kindergarten oral language skill: A key variable in the intergenerational transmission of socioeconomic status. *Research in Social Stratification and Mobility*, 25, 294–305. https://doi.org/10.1016/j.rssm. 2007.03.001

Durkin, K., Simkin, Z., Knox, E and Conti-Ramsden, G. (2009) Specific language impairment and school outcomes. II: Educational context, student satisfaction, and post-compulsory progress, *International Journal of Language & Communication Disorders*, 44:1, 36-55, DOI: 10.1080/13682820801921510

Durkin, K. and Conti-Ramsden, G. (2010) Young people with specific language impairment: A review of social and emotional functioning in adolescence. *Child Language Teaching and Therapy*, 26 (2010), pp. 107-123

Durkin, K., Toseeb, U., Botting, N., Pickles, A., & Conti-Ramsden, G. (2017). Social confidence in early adulthood Forrest et al. 13 among young people with and without a history of language impairment. *Journal of Speech, Language, and Hearing Research*, 60(6), 1635–1647. doi:10.1044/ 2017_JSLHR-L-16-0256

Eadie, P., Morgan, A., Ukoumunne, O. C., Ttofari Eecen, K., Wake, M., & Reilly, S. (2015). Speech sound disorder at 4 years: Prevalence, comorbidities, and predictors in a community cohort of children. *Developmental Medicine & Child Neurology,* 57(6), 578-584.

Early Intervention Foundation (http://earlyinterventionfoundation.org.uk/)

Education Endowment Foundation (2016) https://educationendowmentfoundation.org.uk/news/eef-in-the-news-character-and-education-fund/

Elango S, García JL, Heckman JJ, & Hojman A (2016). Early Childhood Education In Moffitt RA (Ed.), *Economics of Means-Tested Transfer Programs in the United States* (Vol. 2, pp. 235–297). Chicago, IL: University of Chicago Press.

Elliott C, Smith P, McCulloch K. (1997) British Ability Scales. *2nd ed. London, UK: NFER* Nelson;

Hammer CS, Morgan P, Farkas G, et al. (2016): Late Talkers: A Population-Based Study of Risk Factors and School Readiness Consequences. *J Speech Lang Hear Res.* 2017;60(3):607–626. 10.1044/2016_JSLHR-L-15-0417

Erbeli, F. (2019). Translating research findings in genetics of learning disabilities to special education instruction. *Mind, Brain, and Education*, 13(2), 74-79. https://doi.org/10.1111/ mbe.12196

ESRC, Evidence Briefing (2012): Health Inequalities continue to undermine social mobility. http://www.esrc.ac.uk/_images/health-inequalities-undermine-social-mobility_tcm8-20070.pdf

Falk, A.; Kosse, F.; Pinger, P.; Schildberg-Horisch, H. (2016). Take me on Your Shoulders! The Effect of Child Mentoring on Education Outcomes. I*nstEAD - Workshop*. *The role of "soft skills" in life outcomes of children and young people. 14th June 2016.*

Farkas, G., Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race*. Social Science Research, 33,* 464-497.

Feinstein, L. (2003). Inequality in the Early Cognitive Development of British Children in the 1970 Cohort. *Economica*, 70, pp. 73-97.

Felsenfeld, S., Broen, P.A. and McGue, M. (1994). A 28-year follow up of adults with a history of moderate phonological disorder: linguistic and personality results. Journal of Speech and Hearing Research. 35, 1114-1125

Field, F. (2010). The Foundation Years: preventing poor children becoming poor adults - The report of the Independent Review on Poverty and Life Chances. *HM Government*.

Forrest, C. L., Gibson, J. L., Halligan, S. L., & St Clair, M. C. (2018).A longitudinal analysis of early language difficulty and peer problems on later emotional difficulties in adolescence: Evidence from the millennium cohort study. *Autism and Developmental Language Impairments,* 3, 1–15

Frattali C. (1998) Measuring Outcomes in Speech–Language Pathology. *New York: Thieme-Stratton Corp*.

Frizelle, P., O'Toole, C., Lyons, R., Tangney, A., & Murphy, C. A. (2019). Ireland: history of language impairment/developmental language disorder in the Republic of Ireland. In J. Law, C. McKean, C. A. Murphy, & E. T. Thordardottir (Eds.), *Managing Children with Developmental Language Disorder: Theory and Practice across Europe and Beyond* (pp. 272–284). Routledge.

Fu, C. and Wolpin, K. (2013) Structural Estimation of a Becker-Ehrlich Equilibrium Model of Crime: Allocating Police Across Cities to Reduce Crime, *mimeo UPenn.*

Fujiki, B. Brinton, M. Morgan, C.H. Hart (1999)*. Withdrawn and sociable behaviour of children with language impairment.* Language, Speech and Hearing Services in Schools, 30 (1999), pp. 183–195

Fujita Y, Nakamura Y, Hiraoka J, Sakata K (1995) Physical-strength tests and mortality among visitors to health-promotion centers in Japan, *Journal of Clinical Epidemiology,* vol. 48 (pg 1349-59)

Fyfe, E. R., Matz, L. E., Hunt, K. M., & Alibali, M. W. (2019). Mathematical thinking in children with developmental language disorder: The roles of pattern skills and verbal working memory. *Journal of Communication Disorders, 77*, 17–30.

Gale C, Martyn C, Cooper C and Sayer AA (2007) Grip strength, body composition, and mortality. *International Journal of Epidemiology*, Volume 36, Issue 1, February 2007, Pages 228–235

Garcia, J. L., Heckman, J., Leaf, D., and Prados, M. (2017). Quantifying the Life-Cycle Benefits of a Prototypical Early Childhood Program. *IZA DP No. 10811.*

Gerrits, E., Jong, J., Zwitserlood, R., & Klatte, I. (2019). The Netherlands: history of language impairment/developmental language disorder in the Netherlands. In J. Law, C. McKean, C. A. Murphy, & E. T. Thordardottir (Eds.), *Managing Children with Developmental Language Disorder: Theory and Practice across Europe and Beyond* (pp. 339–350). Routledge.

Ghassabian, A., Rescorla, L., Henrichs, J., Jaddoe, V. W., Verhulst, F. C., and Tiemeier, H. (2014) 'Early lexical development and risk of verbal and nonverbal cognitive delay at school age', *Acta Paediatrica, International Journal of Paediatrics*, 103 (1), pp. 70–80.

Girard, L. C., Pingault, J. B., Doyle, O., Falissard, B. and Tremblay, R. E. (2015) 'Developmental Associations Between Conduct Problems and Expressive Language in Early Childhood: A Population-Based Study'*, Journal of Abnormal Child Psychology.*

Gini G., and Pozzoli, T. (2009) Association between bullying and psychosomatic problems: a meta-analysis. *Pediatrics, 123*(3), 1059–1065.

Godfrey Thompson Unit (1978) Edinburgh Reading Test. *Sevenoaks: Hodder and Stoughton*.

Goodman, A., Joshi, H., Nasim, B. and Tyler, C. (2015) Social and Emotional Skills in Childhood and Their Long-Term Effects on Adult Life*. Institute of Education, London*

Goh, S. and O'Kearney, R. (2012) Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairment: Meta-analyses of controlled prospective studies. *Journal of Child Psychology and Psychiatry*. (2012) http://dx.doi.org/10.1111/jcpp.12009

Goldberg D, Williams P. (1991) A user's guide to the general health questionnaire. *London: Nfer-Nelson.*

Goldberg DP, Hillier VF. (1979) A scaled version of the general health questionnaire. *Psychol Med.* 9(1):139–45.

Goldthorpe, J. (2012): Understanding – And Misunderstanding – Social Mobility in Britain: The entry of the Economists, the Confusion of Politicians and the Limits of Educational Policy. *Barnett Papers in Social Research*, Volume 2.

Gottschalk, P. and Smeeding, T. M. (1997) Cross-National Comparisons of Earnings and Income Inequality. *Journal of Economic Literature*, 633-87.

Gough P, and Tunmer W. (1986)Decoding, reading and reading disability. *Remedial and Special Education*;7(1):6–10

Gov.uk (2016) Funding Boost For Schools Helping Pupils Develop Character. https://www.gov.uk/government/news/funding-boost-for-schools-helping-pupils-develop-character Last accessed 06/12/2018

Grawe, N. D. (2006) 'Lifecycle Bias in Estimates of Intergenerational Earnings Persistence', *Labour Economics*, 13(5), 551–570.

Great Ormond Street Hospital (2018) Speech and Language Therapy Clinical Outcomes https://www.gosh.nhs.uk/health-professionals/clinical-outcomes/speech-and-language-therapy-clinical-outcomes *(Last Accessed 3rd June 2019)*

Green, H., McGinty A., Meltzer H., Ford T., and Goodman R. (2005) Mental Health of Children and Young People in Great Britain. *Office of National Statistics.*

Greenwood, C. R. Schnitz, A. G., JudiCarta, J.J., Wallisch, A. & Irvin, D.W. (2020) A systematic review of language intervention research with low-income families: A word gap prevention perspective, *Early Childhood Research Quarterly, 50, 230-45, https://doi.org/10.1016/j.ecresq.2019.04.001.*

Gregg, P; Macmillan, L; Vittori, C (2014): Moving towards estimation lifetime intergenerational economic mobility in the UK http://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/wp332.pdf

Gregory, J. and Bryan, K. (2011) Speech and language therapy intervention with a group of persistent and prolific young offenders in a non-custodial setting with previously undiagnosed speech, language and communication difficulties. *International Journal of Language & Communication Disorders*, 46 (2), pp. 202-215

Griliches, Z and Mason, W. (1972) "Education, Income, and Ability," *Journal of Political Economy* 80, no. 3, Part 2: S74-S103.

The Guardian (2018) Damian Hinds Focuses on Soft Skills in his First Big Speech as Education Secretary https://www.theguardian.com/politics/2018/jan/22/damian-hinds-education-secretary-focuses-soft-skills-first-big-speech?] Last accessed 01/12/2018

Gutman, L., & Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people. A literature review. *London*.

Hahn, J., Todd, P. and Van der Klaauw, W. (1999) "Evaluating the Effect of an Antidiscrimination Law Using a Regression-Discontinuity Design," *NBER Working Papers 7131, National Bureau of Economic Research*, Inc.

Hart, B. and Risley, T.R. (1995) Meaningful differences in the everyday experience of young American children. *Paul Brookes.*

Harrison LJ, 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. *J Speech Lang Hear Res*.53(2):508–529pmid:19786704

Harvey, H. A., and Miller, G. E. (2017). Executive function skills, early mathematics, and vocabulary in Head Start preschool children. *Early Educ. Dev.* 28, 290–307. doi: 10.1080/10409289.2016.1218728

Hassan ES, Darweesh AE, Ibrahim RA, Zareh WM. (2020) Psychological status of school-aged children and adolescents with dyslalia. *J Curr Med Res Pract*;5:217-24

Hattie J. (2009) Visible learning: A synthesis of over 800 meta-analyses relating to achievement. New York: *Routledge*.

Hayes, A. (2009) Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium, *Communication Monographs*, 76:4, 408-420, DOI: 10.1080/03637750903310360

Heberlie AE and Carter AS (2015) Cognitive aspects of young children's experience of economic disadvantage. *Psychological Bulletin*, 141, 723–46.

Heckman, J.J., Moon S-H., Pinto R., Savelyev, P., and Yavitz, A. (2010) "Analyzing Social Experiments as Implemented: A Reexamination of the Evidence From the HighScope Perry Preschool Program." Quantitative Economics 1, 1:1–46. First draft, September, 2006.

Heckman, J.J., Kautz, T., (2013). Fostering and measuring skills: Interventions that improve character and cognition. *In: Heckman JJ, Humphries JE & Kautz T (eds.) The Myth of Achievement Tests: The GED and the Role of Character in American Life*. The University of Chicago Press, Chicago, IL, pp. 341-430

Heckman, J. and Mosso, S. (2014) 'The Economics of Human Development and Social Mobility, Annual Review of Economics', *Annual Reviews*, 6(1): 689-733.

Heckman, J.J.,Humphries J.E., Urz´ua S., and Veramendi G. (2011) "The Effects of Educational Choices on Labor Market, Health, and Social Outcomes." *Unpublished manuscript,* University of Chicago, Department of Economics.

Heckman, J.J., and Masterov D.V. (2007) "The Productivity Argument for Investing in Young Children." *Review of Agricultural Economics* 29, 3: 446–493

Heckman, J.J., Malofeeva L., Pinto R., and Savelyev P. (2012) "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." Unpublished manuscript, University of Chicago, Department of Economics (first draft, 2008). Under revision, *American Economic Review*.

Heckman, J & Kautz, T (2012): Hard Evidence on Soft Skills. http://ftp.iza.org/dp6580.pdf

Hendricks, A. E., Adlof, S. M., Alonzo, C. N., Fox, A. B., & Hogan, T. P. (2019). Identifying children at risk for developmental language disorder using a brief, whole-classroom screen. *Journal of Speech, Language, and Hearing Research*, 62, 896-908.

Heywood, J., Johnson, J. and Brown, M. (2015) User Guide to the Response and Deaths Datasets. *Centre for Longitudinal Studies*

Hirsch D. (2013) An estimate of the cost of child poverty in 2013. *Joseph Rowntree Foundation*. http://www.cpag.org.uk/sites/default/files/Cost of child poverty research update (2013).pdf

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust, P. K. S., & Suma, K. (2015). The Contribution of Early Communication Quality to Low-Income Children's Language Success. *Psychological Science,* 26(7), 1071–1083. https://doi.org/10.1177/0956797615581493

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74, 1368–1378.

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics*, 19(4), 603-629

Hoff, E., and Tian C. (2005). "Socioeconomic Status and Cultural Influences on Language." *ASHA 2004 Research Symposium* 38 (4): 271–278. doi:10.1016/j.jcomdis.2005.02.003.

Honig, A.S. (2007) Oral Language Development. *Early Child Development and Care*, 177(6&7), 581–61

Huber, P.J., (1967). The behavior of maximum likelihood estimates under non-standard conditions. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, pp. 221–233

Hyde, M., Wiggins, R.D., Higgs, P. and Blane, D.B. (2003) A measure of quality of life in early old age: The theory, development and properties of a needs satisfaction model (CASP-19). *Aging & Mental Health*, 7, 186-194.

Ibanez G, Bernard JY, Rondet C, Peyre H, Forhan A, Kaminski M, Saurel-Cubizolles, M.J. (2015) et al. Effects of antenatal maternal depression and anxiety on children's early cognitive development: a prospective cohort study. *PLOS One*; 10: e0135849

I Can, (2017): http://www.ican.org.uk/en/sitecore/content/ICAN2/Global/Components/Rendering%20Elements/Press%20Releases/ONE%20IN%20TEN%20YOUNG%20PEOPLE%20LACK%20LANGUA

GE%20AND%20COMMUNICATION%20SKILLS%20NEEDED%20FOR%20ADULTHOOD%20WARNS%20NEW%20REPORT.aspx Last accessed 23/05/2017

Im-Bolterm, N., & Cohen, N. J. (2007). Language impairment and psychiatric comorbidities language, communication, and literacy: Pathologies and treatments. *Pediatric Clinics of North America*, 54, 525–542.

Im-Bolter, N.J. Cohen, F. Farnia (2013). I thought we were good: Social cognition, figurative language, and psychopathology*. Journal of Child Psychology and Psychiatry*, 54 pp. 724–732

Ingram, T. T. S. (1959) Specific developmental disorders of speech in childhood. *Brain,* 82(3), 450– 467.

Jacob, B.A. (2002). "Where the Boys Aren't: Non-cognitive Skills, Returns to School and the Gender Gap in Higher Education." *Economics of Education Review*. December, 21:6, pp. 589–98.

Jakobsen, J.C., Gluud, C., Wetterslev, J., Winkel, P. (2017) When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. *BMC Med Res Methodol* 17, 162. https://doi.org/10.1186/s12874-017-0442-1

Javorsky, J. (1996). An examination of youth with attention-deficit/hyperactivity disorder and language learning disabilities: A clinical study. *Journal of Learning Disabilities*, 29, 247-258

Jenkins, S. P. (1995) .Did the Middle Class Shrink During the 1980's?: U.K. Evidence from Kernel Density Estimates. *Working Papers of the ESRC Research Centre on Micro Social Change*, University of Essex, March 1995.

Jerrim, J., & Vignoles, A. (2013). Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes*. Journal of the Royal Statistical Society: Series A (Statistics in Society),* 176(4), 887–906.

John, O. and Srivastava, S. (1999) "The Big Five Trait Taxonomy: History, Measurement and Theoretical Perspectives." In *Handbook of Personality: Theory and Research, edited by L. A. Pervin, and O. P. John,* New York: The Guilford Press, chapter 4, 102–138.

Johnson, C; Beitchman, J; Young, A; Escobar, M; Atkinson, L; Wilson, B (1999). Fourteen-Year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language and Hearing Research*. Volume 42 (1) p744-760

Johnson, C.J., Beitchman, J.H. and Brownlie, E.B. (2010). Twenty-year follow-up of children with and without speech language impairments: Family, educational, occupational, and quality of life outcomes. *American Journal of Speech Language Pathology*, 19, 51–65.

Johnson, P., and Kossykh, Y. (2008). Early years, life chances and equality: a literature review. *Equality and Human Rights Commiss*ion. http://cdn.basw.co.uk/upload/basw_25807-5.pdf

Jones, F.L., Kelley, J., (1984). Decomposing Differences between Groups: A Cautionary Note on Measuring Discrimination. *Sociological Methods & Research* 12, 323-343

Jones, A; Wildman, J (2008). Health, income and relative deprivation: Evidence from the BHPS. *Journal of Health Economics*. Volume 27 (2) p308-324

Jongsma, E (1982) Test Review: Peabody Picture Vocabulary Test: Revised (PPVT-R) *Journal of Reading.* Vol. 25, No. 4 (Jan., 1982), pp. 360-364 (5 pages) https://www.jstor.org/stable/40030386

Kautz, T. (2015). Using School Administrative Data to Measure Non-Cognitive Skills. *HCEO Conference on Measuring and Assessing Skills* October 2, 2015. https://cehd.uchicago.edu/wp-content/uploads/2017/01/Kautz.pdf

Kerckhoff, A., Ezell, E., Brown, J. (2002) Toward an improved measure of educational attainment in social stratification research. *Soc. Sci. Res*., 31, pp. 99-123 https://doi.org/10.1006/ssre.2001.0720

Kline, P. and Walters, C. (2014). “Evaluating Public Programs with Close Substitutes: The Case of Head Start”. *IRLE Working Paper No. 123-14*. http://irle.berkeley.edu/workingpapers/123-14.pdf

Kraljević, J. K., Dokoza, K. P., & Matić, A. (2019). Croatia: history of language impairment/developmental language disorder in Croatia. In J. Law, C. McKean, C. A. Murphy,

& E. T. Thordardottir (Eds.), *Managing Children with Developmental Language Disorder: Theory and Practice across Europe and Beyond* (pp. 169–178). Routledge.

Kunnari, S., Laasonen, M., & Smolander, S. (2019). Finland: history of language impairment/developmental language disorder in Finland. In J. Law, C. McKean, C. A. Murphy, & E. T. Thordardottir (Eds.), *Managing Children with Developmental Language Disorder: Theory and Practice across Europe and Beyond* (pp. 215–224). Routledge.

Law, J., Boyle, J., Harris, F., Harkness, A. and 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 and Communication Disorders*, 35 (2), pp. 165–88.

Law, J., Garrett, Z., & Nye, C. (2004). The efficacy of treatment for children with developmental speech and language delay/ disorder: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 47, 924–943.

Law,J. Rush, R, Parsons, S. & Schoon, I. (2009) Modelling developmental language difficulties from school entry into adulthood: Literacy, mental health and employment outcomes. *Journal of Speech, Language and Hearing Research* 52, 1401-1416.

Law, J. and Elliott, L., (2009). The relationship between communication and behaviour in children: a case for public mental health. *The Journal of Public Mental Health*, 8(1); 4-11.

Law, J., McBean, K. and 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, 657–664

Law, J., Rush, R., Anandan, C., Cox, M. and Wood, R. (2012a) 'Predicting language change between 3 and 5 Years and its implications for early identification', *Pediatrics,* 130 (1), pp. 132–37.

Law, J; Lee, W; Roulstone, S; Wren, Y; Zeng, B; Lindsay, G (2012b): What works: Interventions for children and young people with speech language and communication needs.

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/219623/DFE-RR247-BCRP10.pdf Nottingham: DfE.

Law, J., Rush, R., Parsons, S., & Schoon, I. (2013). The relationship between gender, receptive vocabulary, and literacy from school entry through to adulthood. *Int J Speech Lang Pathol*, 15(4), 407-415. doi:10.3109/17549507.2012.721897

Law, J., Charlton, J., Dockrell, J., Gascoigne, M., McKean, C. and Theakston, A. (2017a) Early Language Development: Needs, Provision, and Intervention for Preschool Children from Socioeconomically Disadvantaged Backgrounds . *Report for the Education Endowment Foundation,* October (Education Endowment Foundation). https://educationendowmentfoundation.org.uk/public/files/Law_et_al_Early_Language_Development_final.pdf

Law J, Levickis P, McKean C, Goldfeld S, Snow P and Reilly S. (2017b) Child Language in a Public Health Context*. Melbourne: Murdoch Childrens Research Institute*.

Law, J., Charlton, J. and Dennis, J. (2017c*) Cochrane Protocol B9805a:* Speech and language therapy interventions for children with primary speech and/or language disorders *Cochrane Library* http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD012490/full

Law, J., McKean, C., Murphy, A. and Thordardottir, A.T. (2019a) (Eds.), Managing Children with Developmental Language Disorder: Theory and Practice across Europe and Beyond

Law, J. (2019b). The changing perception of communication needs—a litmus Test for the Warnock legacy. Front. Educ. 4:42. doi: 10.3389/feduc.2019.00042

Lazear, E. (1981). Agency, Earnings Profiles, Productivity, and Hours Restrictions. *The American Economic Review, 71*(4), 606-620. Retrieved July 3, 2020, from www.jstor.org/stable/1806184

Le, H.N., Mensah, F., Eadie, P., McKean, C., Sciberras, E., Bavin, E.L., Reilly, S. and Gold, L. (2020a), Health-related quality of life of children with low language from early childhood to adolescence: results from an Australian longitudinal population-based study. *J Child Psychol Psychiatr*. doi:10.1111/jcpp.13277

Le, H.N., Le, L.K, Nguyen, P., Mudiyanselage, S., Eadie, P., Mensah, F., Sciberras, E., Gold, L. (2020b) Health-related quality of life, service utilization and costs of low language: A systematic review *International Journal of Language and Communication Disorders*, VOL. 55, NO. 1, 3–25

Lee, D. and Card, D. (2008) "Regression discontinuity inference with specification error," *Journal of Econometrics*, 142 (2), 655 – 674. The regression discontinuity design: Theory and applications.

Lee, D. and Lemieux, T. (2010) "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48 (2), 281–355.

Leonard LB. (2014) Children with Specific Language Impairment. *Cambridge: MIT Press.*

Letts, C., Edwards, S., Sinka, I., Schaefer, B. and Gibbons, W., (2013), Socio-economic status and language acquisition: children's performance on the new Reynell Developmental Language Scales. *International Journal of Language and Communication Disorders*, 48, 131–143.

Levickis, P., Reilly, S., Girolametto, L., Ukoumunne, O. C., and Wake, M. (2018a). Associations between maternal responsive linguistic input and child language performance at age 4 in a community-based sample of slow-to-talk toddlers. *Child Care Health Dev.* 44, 776–783. doi: 10.1111/cch.12600

Levickis, P., Sciberras, E., McKean, C., Conway, L., Pezic, A., Mensah, F., Bavin, E., Bretherton, L., Eadie, P. and Reilly, S. (2018b) Language and social-emotional and behavioural wellbeing from 4 to 7 years: a community-based study. *Eur Child Adolesc Psychiatry* 27, 849–859 https://doi.org/10.1007/s00787-017-1079-7

Lewis, E (2011). Immigrant-Native Substitutability: The Role of Language Ability. http://www.nber.org/papers/w17609

Lindqvist, E.; Westman, R. (2009) : The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment*, IFN Working Paper*, No. 794, Research Institute of Industrial Economics (IFN), Stockholm

Lindsay, G; Dockrell, J; Law, J; Roulstone, S (2012). *The Better Communication Research Programme: Improving Provision for children and young people with speech, language and communication needs* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/219622/DFE-RR247-BCRP1.pdf

Lindsay, G and Dockrell, J (2012). Longitudinal patterns of behavioural, emotional and social difficulties and self-concepts in adolescents with a history of specific language impairment. *Language Speech and Hearing Services in Schools*, 43(4) pp, 445-460

Lindsay, G., & Strand, S. (2016). Children with Language Impairment: Prevalence, Associated Difficulties, and Ethnic Disproportionality in an English Population. *Frontiers in Education*, 1(2), 1-14. doi:10.3389/feduc.2016.00002.

Locke, A., Ginsborg, J., & Peers, I. (2002). Development and disadvantage: Implications for the early years and beyond. *International Journal of Language and Communication Disorders*. https://doi.org/10.1080/13682820110089911

Löhndorf, R. T., & Mostert & Van Onderen. (2018). Parenting and professional childcare in Chile : relations with child developmental outcomes in preschoolers. https://openaccess.leidenuniv.nl/handle/1887/60213

Longfield, A. (2019) We Need To Talk: Access to Speech and Language Therapy, *Children's Commissioner,* June 2019: Accessible at https://www.childrenscommissioner.gov.uk/wp-content/uploads/2019/06/cco-we-need-to-talk-june-2019.pdf *(Last Accessed: 29/08/2020)*

Lowden, K. Hall, S et al. (2011) Employers' perceptions of the employability skills of new graduates. *Glasgow: University of Glasgow SCRE Centre and Edge Foundation*

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) '''general intelligence,' objectively determined and measured.'' *Journal of Personality and Social Psychology*, 78, 96–111

Lutz W, Skirbekk V, Testa MR. The low fertility trap hypothesis: forces that may lead to further postponement and fewer births in Europe. *Vienna Yearbook of Population Research*. Vienna: Vienna Institute of Demography; 2006. pp. 167–192.

Mangalore, R. (2006) Income, health and health care utilization in the UK. *Applied Economics*. 38(6): p. 605-17.

Manso, J. M. M., García-Baamonde, M. E., & Alonso, M. B. (2011). Design of a language stimulation program for children suffering abuse. *Children and Youth Services Review*, *33*(7), 1325–1331. https://doi.org/https://doi.org/10.1016/j.childyouth.2011.03.003

Markham, C., Van Laar, D., Gibbard, D. and Dean, T. (2009). Children with speech, language and communication needs: their perceptions of their quality of life. *International Journal of Language & Communication Disorders, 44,* 748-768.

Marmot, M., Allen, J., Goldblatt, P. Boyce, T. McNeish, D. Grady, M. and Geddes, I. (2010). Fair society, healthy lives. *Institute for Health Equity*, University College London http://www.instituteofhealthequity.org/resources-reports/fair-society-healthy-lives-the-marmot-review (visited on 22/09/2020).

Marmot M, Allen J, Boyce T, Goldblatt P, Morrison J. (2020) Health equity in England: the Marmot review 10 years on.*Institute of Health Equity*.

Marsh, K., Bertranou, E., Suominen, H., & Venkatachalam, M. (2010). An economic evaluation of speech and language therapy. p1–90. http://www.optimityadvisors.com/economic-evaluation-speech-and-language-therapy

Masters, S.H. (1974) "The Effect of Educational Differences and Labor Market Discrimination on the Relative Earnings of Black Men," *Journal of Human Resources* 9, 342-360.

McCauley R. J., & Strand E. A. (2009). A review of standardized tests of nonverbal oral and speech motor performance in children. *American Journal of Speech-Language Pathology*, 17, 81–91

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, 43, 947–959

McCrae, R. R., & Costa, P. T. (1987). Validation of the Five-Factor Model of Personality Across Instruments and Observers. *Journal of Personality and Social Psychology*, 52(181–90). https://doi.org/10.1037/0022-3514.52.1.81

McCrary, Justin (2008) “Manipulation of the running variable in the regression discontinuity design: A density test,” *Journal of Econometrics*, 142 (2), 698–714.

McGregor, K. (2020) How we fail children with developmental language disorder*. Language, Speech and Hearing Services in Schools*. Doi:10.1044/2020_LSHSS-20-00003, available at https://pubs.asha.org/doi/abs/10.1044/2020_LSHSS-20-00003 (Last accessed: 25th September 2020).

McKean, C., Wraith, D., Eadie, P., Cook, F., Mensah, F., & Reilly, S. (2017a). Subgroups in language trajectories from 4 to 11 years: The nature and predictors of stable, improving and decreasing language trajectory groups. *Journal of Child Psychology and Psychiatry*, 58, 1081-1091.

McKean, C., Reilly, S., Bavin, E. L., Bretherton, L., Cini, E., Conway, L., Cook, F., Eadie, P., Prior, M., Wake, M. and Mensah, F. (2017b). Language Outcomes at 7 Years: Early Predictors and Co-Occurring Difficulties. *Pediatrics*. doi:10.1542/peds.2016-1684

McLeod, S., Harrison, L. J., & Wang, C. (2019). A longitudinal population study of literacy and numeracy outcomes for children identified with speech, language, and communication needs in early childhood. *Early Childhood Research Quarterly*, 47, 507–517. https://doi.org/10.1016/j.ecresq.2018.07.004

Menting, B., van Lier, P.A. and Koot, H.M. (2011), Language skills, peer rejection, and the development of externalizing behavior from kindergarten to fourth grade. *Journal of Child Psychology and Psychiatry*, 52: 72-79. doi:10.1111/j.1469-7610.2010.02279.x

Miller FJW, Court SDM, Walton WS, Knox EG. (1960), Growing up in Newcastle upon Tyne, *London Oxford University Press*

Miller F.J.W., Court S.D.M., Knox E.G. , Brandon S. (1974), The School Years in Newcastle upon Tyne, 1952-62 *Oxford University Press (London)*

Mincer, J. (1962): “On-the-Job Training: Costs, Returns, and Some Implications,” *The Journal of Political Economy*, 70(5), 50—79.

Mok PL, Pickles A, Durkin K, Conti-Ramsden G (2014) Longitudinal trajectories of peer relations in children with specific language impairment. *Journal of Child Psychology and Psychiatry* 55(5):516–527

Mol, S. E. and Bus, A. G. (2011) 'To Read or Not to Read: A Meta-Analysis of Print Exposure From Infancy to Early Adulthood'*, Psychological Bulletin*, 137 (2), pp. 267–96.

Morgan, A, Eecen, K.T., Pezic, A., Brommeyer, K., Mei, C., Eadie, P., Reilly, S., and Dodd, B. (2017) "Who to refer for speech therapy at 4 years of age versus who to "watch and wait"?." *The Journal of pediatrics* 185: 200-204.

Morley, M., (1965) The Development and Disorders of Speech in Childhood *(Edinburgh: E&S Livingstone).*

Morley, M., Court, D., and Miller, H. (1954). *British Medical Journal*, 1, 8.

Morley, M., Court, D., Miller, H., Garside, R. F. (1955) 'Delayed speech and developmental aphasia.' *British Medical Journal* ii, 463.

Mullen R. (2004) Evidence for whom?: ASHA's National Outcomes Measurement System. *Journal of Communication Disorders* 37:413–17. 10.1016/j.jcomdis.2004.04.004.

Nathan, L., Stackhouse, J., Goulandris, N., & Snowling, M. J. (2004). Educational consequences of developmental speech disorder: Key stage 1 National Curriculum assessment results in English and mathematics*. British Journal of Educational Psychology*, 74, 173–186.

National Literacy Trust (2018). Home Matters: Making the most of the home environment*.* Retrieved from *https://cdn.literacytrust.org.uk/media/documents/Home_matters_-_making_the_most_of_the_home_learning_environment.pdf* Last accessed 15/11/2018

National Statistics (2018): Early years foundation stage profile (EYFSP) results: 2017 to 2018 Additional Tables, Table 1. Available at: https://www.gov.uk/government/statistics/early-years-foundation-stage-profile-results-2017-to-2018

National Statistics (2019) Key Stage 4 and Multi-Academy Trust Performance 2018 (revised). Available at: https://www.gov.uk/government/statistics/key-stage-4-and-multi-academy-trust-performance-2018-revised *Last accessed 24/08/2020*

Nelson, R. J., Benner, G. J., Stern, N., & Stage, S. A. (2006). Interrelationships among language skills, externalizing behavior, and academic fluency and their impact on the academic skills on students with ED. *Journal of Emotional and Behavioral Disorders*, 14, 209–216.

Neuman S.B., Dickinson D.K., editors. (2011) Handbook of early literacy research. *New York: The Guilford Press;*

Nickisch, A., & von Kries, R. (2009). Short-Term Memory (STM) Constraints in Children With Specific Language Impairment (SLI): Are There Differences Between Receptive and Expressive SLI? *Journal of Speech Language and Hearing Research*, *52*(3), 578. https://doi.org/10.1044/1092-4388(2008/07-0150

Nitido, H. and Plante, E. (2020) Diagnosis of Developmental Language Disorder in Research Studies. *Journal of Speech, Language and Hearing Research*. Research Note. Vol. 63 (8) August 2020 p2777-2788. https://doi.org/10.1044/2020_JSLHR-20-00091

Nguyen, H.T., Connelly, L., Le, H.T., Mitrou, F., Taylor, C., Zubrick, S., (2019). Sources of Ethnicity Differences in Non-Cognitive Development in Children and Adolescents. *Life Course Centre Working Paper Series, 2019-21. Institute for Social Science Research*, The University of Queensland

Norbury, C.F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G. & 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*, 57, 1247–1257.

Osborne-Groves, M. (2004, March). How Important is Your Personality? Labor Market Returns to Personality for Women in the US and UK. Unpublished manuscript, Towson University, Department of Economics. In press at *Journal of Economic Psychology*.

Parsons, S. (2014) 'Childhood cognition in the 1970 British Cohort Study' *Data note*

Parsons, S., Schoon, I., Rush, R. and Law, J. (2011), Long-term outcomes for children with early language problems: beating the odds. *Children and Society*, 25, 202–214.

Pearce M.S., Deary I.J., Young A.H., Parker, L. (2005) Growth in early life and childhood IQ at age 11 years: the Newcastle Thousand Families Study. *International Journal of Epidemiology*. 34:673–677

Pearce, M., Unwin, N., Parker, L., Craft, A. (2009) Cohort Profile: The Newcastle Thousand Families 1947 Birth Cohort, *International Journal of Epidemiology*, Volume 38, Issue 4, August 2009, Pages 932–937, https://doi.org/10.1093/ije/dyn184

Peckham, C. S., (1973), Speech defects in a national sample of children aged seven years. *British Journal of Disorders of Communication*, 8, 2–8.

Pianta RC, editors. (2012) Handbook of early childhood education. *New York: The Guilford Press*.

Piketty, T. (2000) 'Theories of Persistent Inequality and Intergenerational Mobility' in A. Atkinson and F. Bourguignon (eds.) Handbook of Income Distribution, Volume 1, *Elsevier* North Holland.

Plante, E. and Vance, R. (1994) Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools* 25, 1524.

Powell & Snow. (2012). Youth (in)justice: oral language competence in early life and risk for engagement in anti-social behaviour in adolescence.

Prevoo, T. and Weel, B. (2012) The importance of early conscientiousness for socio-economic outcomes: evidence from the British Cohort Study. *Oxford Economic Papers*, 2015, 1–31 doi: 10.1093/oep/gpv022

Psacharopoulos, G. and H.A. Patrinos. (2018). "Returns to Investment in Education: A Decennial Review of the Global Literature." *Education Economic*, 26 (5): .445-458.

Raven, J. C. (1958). Guide to using the Mill Hill Vocabulary Scale with the Progressive Matrices Scales. *H. K. Lewis & Co.*

RCSLT (2016) NEF Consulting. *Stoke Speaks Out campaign.* Retrieved from: https://www.rcslt.org/news/news/2016_news_archive/stoke_speaks_out. Last accessed 15/11/2018

Records, N. L., Tomblin, J.B. and Freese, P.R. (1992). The quality of life of young adults with histories of specific language impairment. *American Journal of Speech and Language Pathology*, 1, 44-53

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., Conway, L., Eadie, P., and Bretherton, L. (2010) 'Predicting language outcomes at 4 years of age: Findings from the Early Language in Victoria Study', *Pediatrics,* 126 (6), pp. e1530–37.

Reilly, S., McKean, C., Levickis, P. (2014) Late talking: can it predict later language difficulties? *Centre for Research Excellence in Child Language, Research Snapshot 2.*

Roben, C., Cole, K. P., & Armstrong, L. M. (2013). Longitudinal relations among language skills, anger expression, and regulatory strategies in early childhood. *Child Development*, 84, 891–905. https://doi.org/10.1111/cdev.12027

Roberts, B. W., N. R. Kuncel, R. L. Shiner, A. Caspi, and L. R. Goldberg. (2007) "The power of personality: The comparative validity of personality traits, socioeconomic status, and 49 cognitive ability for predicting important life outcomes." *Perspectives in Psychological Science* 2, 4: 313–345.

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. https://doi.org/10.1177/0956797617742725

Rose, E., Lehrl, S., Ebert., S and Weinert, S. (2018) Long-Term Relations Between Children's Language, the Home Literacy Environment, and Socioemotional Development From Ages 3 to 8, Early Education and Development, 29:3, 342-356, DOI: 10.1080/10409289.2017.1409096

Roulstone, S., Law, J., Rush, R., Clegg, J. and Peters, T. (2011) Investigating the Role of Language in Children's Early Educational Outcomes *(Bristol: University of the West of England).*

Roulstone, S., Wren, Y., Bakopoulou, I., Goodlad, S., & Lindsay, G. (2012). Exploring Interventions For Children And Young People With Speech, Language And Communication Needs: A Study Of Practice. *Better Communication Research Programme*, 86. http://dera.ioe.ac.uk/16323/1/DFE-RR247-BCRP10a.pdf

Roulstone SE, Marshall JE, Powell GG, et al. (2015) Evidence-based intervention for preschool children with primary speech and language impairments: Child Talk – an exploratory mixed-methods study. Southampton (UK*): NIHR Journals Library; (Programme Grants for Applied Research, No. 3.5.)* Chapter 5, Identifying a toolkit for assessment and outcome measurement. Available from: https://www.ncbi.nlm.nih.gov/books/NBK311158/

Rowe, M. L. (2008). "Child-Directed Speech: Relation to Socioeconomic Status, Knowledge of Child Development and Child Vocabulary Skill." *Journal of Child Language* 35 (1): 185–205. doi:10.1017/S0305000907008343.

Sampson, G. (2005). The 'Language Instinct' Debate. *London: Continuum*

Save The Children (2014) *Born to Read – Read on. Get on.* Clark, C; Morrisroe, J; Law, J; King, T.

Sawilowsky, S. S. (2009). New effect size rules of thumb*. Journal of Modern Applied Statistical Methods*, 8(2), 597 – 599.

Schonell, F. (1942). Backwardness in the basic subjects. *London, UK: Oliver & Boyd*

Schoon, I., Parsons, S. Rush, R., & Law,J., (2010a) Children's language and ability and literacy development: A twenty-nine year follow-up study. *Pediatrics.* Published online Feb 8, 2010; DOI: 10.1542/peds.2008-2111

Schoon, I., Parsons, S. Rush, R., & Law, J., (2010b) Children's language ability and psychosocial development:A twenty-nine year follow-up study. *Pediatrics. 126, e73-e80*

Schulz, K. F., Altman, D. G., Moher, D., Barbour, V., Berlin, J. A., Boutron, I., Devereaux, P. J., Dickersin, K., Elbourne, D., Ellenberg, S., Gebski, V., Goodman, S., Gøtzsche, P. C., Groves, T., Grunberg, S., Haynes, B., Hopewell, S., James, A., Juhn, P., Guyatt, G. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *Ann Int Med*, *152*. https://doi.org/10.3736/jcim20100702

Schulz B. (2008). The importance of soft skills: education beyond academic knowledge. NAWA *Journal of Language and Communication* p146-154. http://ir.polytechnic.edu.na/bitstream/10628/39/1/TheImportanceofSoftSkills-Educationbeyondacademicknowledge.pdf

Scott-Phillips T. C. and Blythe R. A., (2013) Why is combinatorial communication rare in the natural world, and why is language an exception to this trend? *J. R. Soc. Interface.*1020130520http://doi.org/10.1098/rsif.2013.0520

Sénéchal, M., & LeFevre, J. -A. (2001). Storybook reading and parent teaching: Links to language and literacy development. *In: P. R. Britto & J. Brooks-Gunn (Eds.),* The role of family literacy environments in promoting young children's emerging literacy skills (pp. 39–52). *San Francisco: Jossey–Bass.*

Shonkoff, J., & Phillips, D. (2000). From neurons to neighborhoods : The science of early child development. *Washington, D.C.: National Academy Press.*

Shriberg, L. D. (2010) Childhood speech sound disorders: from postbehavorism to the postgenomic era. *In R. Paul and P. Flipsen (eds), Speech Sound Disorders in Children—in Honor of Lawrence D. Shriberg (San Diego, CA: Plural),* pp. 1– 34.

Singh, S. and Maddala, G. (1976) 'A Function for Size Distribution of Incomes' *Econometrica,* Vol. 44, pp. 963-970.

Snow, P. C. (2020). SOLAR: The Science of Language and Reading. *Child Language Teaching and Therapy*. https://doi.org/10.1177/0265659020947817

Snowling, M.J., Adams, J.W., Bishop, D.V.M., Chipchase, B.B. and Kaplan, C. (2001). Educational attainments of school leavers with a preschool history of speech-language impairments*. International Journal of Language and Communication Disorders*, 36, 173-183.

Snowling MJ, Duff FJ, Nash HM, Hulme C. (2016) Language profiles and literacy outcomes of children with resolving, emerging, or persisting language impairments. *Journal of Child Psychology and Psychiatry*. ;57(12):1360-1369. doi:10.1111/jcpp.12497

Sobel, M. E. (1982). Aysmptotic confidence intervals for indirect effects in structural equation models*. In S. Leinhardt (Ed.), Sociological methodology (pp. 290212). San Francisco: Jossey-Bass*

Solon, G. (1999) Intergenerational Mobility in the Labor Market, Chapter 29 *in O. Ashenfelter and D. Card (eds.) Handbook of Labor Economics, North Holland Press.*

Spaulding TJ, Plante E, Farinella KA (2006) Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech and Hearing Services in Schools* 37: 61–72. pmid:16615750

Spence J, Walton WS, Miller FJW, Court SDM. *(1954) A Thousand Families in Newcastle upon Tyne, London, Oxford University Press*

St Clair, M.C., A. Pickles, K. Durkin, G. Conti-Ramsden (2011) A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of specific language impairment (SLI). *Journal of Communication Disorders*, 44, pp. 186–199

Stockman, I. (2000) The New Peabody Picture Vocabulary Test-III: An Illusion of Unbiased Assessment?. Available from: https://www.researchgate.net/publication/286955330_The_New_Peabody_Picture_Vocabulary_Test-III_An_Illusion_of_Unbiased_Assessment [accessed Mar 01 2019].

Tomblin, J. B. Smith, E. and Zhang, X., (1997), Epidemiology of specific language impairment: prenatal and peri-natal risk factors. *Journal of Communication Disorders*, 30, 325–344.

Strand, S. & Lindsay, G. (2009) Evidence of ethnic disproportionality in special education in an English population, *Journal of Special Education*, 43(3), 174–190.

Strand, S. (2014) "Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and 'Getting it'for the white working class." *Research Papers in Education* 29(2):131-71.

Sullivan, A., Moulton, V. and Fitzsimons, E. (2017) The intergenerational transmission of vocabular*y. Centre for Longitudinal Studies Working paper 2017/14*

Taylor, C; Christensen, D; Lawrence, D; Mitrou, F; Zubrick, S (2013) Risk Factors for Children's Receptive Vocabulary Development from Four to Eight Years in the Longitudinal Study of Australian Children; *PLOS-ONE* https://doi.org/10.1371/journal.pone.0073046

Terman, L., Baldwin, B., Bronson, E. et al. (1925) Genetic Studies of Genius: Mental and Physical Traits of a Thousand Gifted Children*, volume 1. Stanford University, CA: Stanford University Press, 1925.*

Tomblin JB, Records NL, 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*. 40:1245–1260. *Folia Phoniatrica et Logopaedica, 67 (2015), pp. 293-299, 10.1159/000444750*

Tomblin, J. B. (2008). Validating diagnostic standards for specific language impairment using adolescent outcomes. In C. F. Norbury, J. B. Tomblin, & D. V. M. Bishop (Eds.), *Understanding developmental language disorders: From theory to practice* (p. 93–114). *Psychology Press.*

Toseeb, U., Pickles, A., Durkin, K., Botting, N. and Conti-Ramsden, G. (2017) Prosociality from early adolescence to young adulthood: A longitudinal study of individuals with a history of language impairment. *Research in Developmental Disabilities*, 62, pp. 148-159,

Toseeb, U., Gibson, J.L., Newbury, D.F., Orlik, W., Durkin, K., Pickles, A. and Conti-Ramsden, G. (2020), Play and prosociality are associated with fewer externalizing problems in children with developmental language disorder: The role of early language and communication environment. *International Journal of Language & Communication Disorders.* doi:10.1111/1460-6984.12541

Toseeb, U and St-Clair, M. C. (2020) Trajectories of prosociality from early to middle childhood in children at risk of Developmental Language Disorder, *Journal of Communication Disorders,* Volume 85, 105984, ISSN 0021-9924, https://doi.org/10.1016/j.jcomdis.2020.105984.

Tucker-Drob EM, Bates TC. (2016) Large cross-national differences in gene× socioeconomic status interaction on intelligence. *Psychological Science*

Ullrich, D., Ullrich, K. and Marten, M., Bronwyn Frances Ewing (Reviewing Editor) (2017) IQ, non-cognitive and social-emotional parameters influencing education in speech- and language-impaired children, *Cogent Education*, 4:1, DOI: 10.1080/2331186X.2017.1357912

Unger JM, Rauch A, Frese M, Rosenbusch N. (2011). Human capital and entrepreneurial success: a meta-analytical review. *J. Bus. Ventur.* 26(3):341–58

Van Agt, H., Verhoeven, L., Van Den Brink, G., &De Koning, H. (2010). The impact on socio-emotional development and quality of life of language impairment in 8-year-old children. *Developmental Medicine and Child Neurology*, 53,81–88

Van Daal, J., Verhoeven, L. and Van Balkom, H. (2007), Behaviour problems in children with language impairment. *Journal of Child Psychology and Psychiatry*, 48: 1139-1147. doi:10.1111/j.1469-7610.2007.01790.x

VanderWeele TJ. (2015) Explanation in Causal Inference: Methods for Mediation and Interaction. *New York, NY: Oxford University Press*.

Villamizar-Villegas, M., Pinzon-Puerto, F., Ruiz-Sanchez, M.A. (2020) A Comprehensive History of Regression Discontinuity Designs: An Empirical Survey of the last 60 Years∗ *Borradores de ECONOMIA.https://repositorio.banrep.gov.co/bitstream/handle/20.500.12134/9832/be_1112.pdf?sequence=1*

Voci, S., Beitchman, J., Brownlie, E., & Wilson, B. (2006). Social anxiety in late adolescence: The importance of early childhood language impairment. *Anxiety Disorders*, 20, 915–930

Wadman R, Durkin K and Conti-Ramsden G (2008) Self-esteem, shyness and sociability in adolescents with specific language impairment. *Journal of Speech, Language and Hearing Research* 51: 938–52.

Wadsworth M, Kuh D, Richards M, Hardy R (2006) Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). *International Journal of Epidemiology* 35:49–54

Walsh, W. B., & Eggerth, D. E. (2005). Vocational Psychology and Personality: The Relationship of the Five-Factor Model to Job Performance and Job Satisfaction. In *Handbook of Vocational Psychology: Theory, Research, and Practice (3rd ed.).*

Wareham, H., Law, J., Huang, W., Labuschagne, L., Attig, M., Franken, M-C. & Weinert, S. (in preparation) The effect of maternal education on children's vocabulary and its impact on social inequality, analysis from three European cohort studies. *Journal of Educational Policy*

Waring, R., & Knight, R. (2013). How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems. *International Journal of Language & Communication Disorders*, 48, 25–40.

Washington, J., & Craig, H. (1992).Performances of low-income, African American preschool and kindergarten children on the Peabody Picture Vocabulary Test–Revised. *Language, Speech, and Hearing Services in Schools*, 23, 329–333

Washington, ,J. A. & Craig, H. K. (1999).Performances of at-risk, African-American preschoolers on the Peabody Picture Vocabulary Test-III. *Language, Speech, and Hearing Services in Schools*, 30, 75–82

Wehner, C., Schils, T. and Borghans, L. (2016) Personality and Mental Health: The Role and Substitution Effect of Emotional Stability and Conscientiousness. *IZA DP No. 10337 Institute for the Study of Labor*

Westrupp, E. M., Reilly, S., McKean, C., Law, J., Mensah, F., & Nicholson, J. M. (2020). Vocabulary development and trajectories of behavioral and emotional difficulties via academic ability and peer problems. *Child Development*. Advance online publication, Jan. 2019. http://dx.doi.org/10 .1111/cdev.13219

White, H., (1982). Maximum likelihood estimation of mis-specified models. *Econometrica* 50, 1–25.

White, IR; Carlin, JB (2010) Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statist.* Med., 29: 2920–2931. doi: 10.1002/sim.3944

Whitehouse, A. J. O., Robinson, M. and Zubrick, S. R. (2011) 'Late talking and the risk for psychosocial problems during childhood and adolescence', *Pediatrics,* 128 (2), pp. e324–32.

Wicherts JM, et al. (2016) Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology* 7:12

Wiig, E. H., Secord, W. and Semel, E., (1992), Clinical Evaluation of Language Fundamentals—Preschool *(London: Psychological Corporation).*

Wildman, J (2003). Income related inequalities in mental health in Great Britain: analysing the causes of health inequality over time. *Journal of Health Economics*. Volume 22 (2) p295-312.

Williams, K & Wang, J (1997) Technical References to the Peabody Picture Vocabulary Test-(PPVT-III). *American Guidance Service.*

Willingham, D. T. (2012). Ask the Cognitive Scientist: Why Does Family Wealth Affect Learning? American Educator, 36(1), 33-3926

Wren, Y., Miller, L. L., Peters, T. J., Emond, A., & Roulstone, S. (2016). Prevalence and predictors of persistent speech sound disorder at eight years old: Findings from a population cohort study. Journal of Speech, Language, and Hearing Research, 59(4), 647-673.

von Hippel, P., Workman, J., & Downey, D. B. (2017). Inequality in reading and math skills comes mainly from early childhood: A replication, and partial correction, of 'Are schools the great equalizer?'. https://dx.d oi.org/10.2139/ssrn.3036094

Yairi, E., & Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. *Journal of fluency disorders*, 38(2), 66–87. https://doi.org/10.1016/j.jfludis.2012.11.002

Yew, S.G.K., O'Kearney, R. (2015) The Role of Early Language Difficulties in the Trajectories of Conduct Problems Across Childhood. *Journal of Abnormal Child Psychology* 43, 1515–1527. https://doi.org/10.1007/s10802-015-0040-9

Zhao, K., Ferguson, E., & Smillie, L. D. (2016). Prosocial Personality Traits Differentially Predict Egalitarianism, Generosity, and Reciprocity in Economic Games. *Frontiers in psychology*, 7, 1137. https://doi.org/10.3389/fpsyg.2016.01137

Zubrick, S. R., Taylor, C. L. and Christensen, D. (2015) 'Patterns and predictors of language and literacy abilities 4–10 years in the longitudinal study of Australian children', *PLoS ONE*, 10 (9).

Zubrick, S. R., Taylor, C. L., Rice, M. L. and 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), pp. 1,562–92.