The influence of child and social factors on the efficacy of language interventions and the role of language in predicting school readiness

Josie Tulip

Thesis submitted for the Degree of Doctor of Philosophy

Faculty of Humanities and Social Science

Education, Communication and Language Sciences

March 2022

Abstract

Background: Poor preschool language and readiness for school can have consequences on life outcomes. Interventions are often utilised to both promote language, and benefit many abilities underpinning school readiness. Intervention implementation and evaluations are commonly motivated by two implicit assumptions: 1) all children will benefit equally from interventions, and 2) children's language gains will benefit school readiness equally. However, language and school readiness are both related to child and family-related social factors through a range of possible mechanisms. Thus, children could be subject to a 'triple threat' of disadvantage – where their developmental and social disadvantages lead to poorer language and school readiness outcomes, poorer intervention response, and less benefit in school readiness from language gains.

Methods: *Phase 1:* a systematic review of language intervention studies examined whether children benefitted equally from interventions, or if gains were affected by child and social factors. *Phase 2:* a secondary data analysis of the Millennium Cohort Study examined if children benefit equally in school readiness from language gains, or if benefits are moderated by child and social factors.

Results: *Phase 1:* Children with more severe language difficulties gained more from interventions in general language, word knowledge, and expressive morphosyntax, but less in listening comprehension. Children with speech difficulties gained less from phonological awareness and expressive morphosyntax interventions. *Phase 2:* Males compared to females, and children living in poverty compared to their more affluent peers benefitted more in school readiness from gains in expressive vocabulary. *Overall:* Being male did not create a 'triple threat' of disadvantage. Speech difficulties created a 'double threat'.

Conclusions: Language, school readiness, child, and social factors may associate with one-another through complex mechanisms which are not just based on additive risk. This has implications on how interventions targeting language and school readiness are assessed and implemented, and so requires further investigation.

Dedication

There are a number of people who helped me through completing this thesis.

Cristina, it has been an absolute privilege to have you as my supervisor. Your vast knowledge and unconditional support of me in all things has helped me develop into someone I can be proud of both professionally and personally. Thank you for all the pep talks, kind ears and laughs.

Carolyn, I have really appreciated you stepping up to be my supervisor in the last few months of my PhD. Your perspectives have been invaluable to me, and your kind encouragement and thorough feedback really helped over this time.

James, I am so sad that you could not see the day I finished this. But if in some way you are able to read this, I just want you to know that I will always cherish what you have done for me over the 6 years that I knew you. I will make sure to bake some brownies and have a whiskey in your honour when this is submitted.

Spencer, without your love and support I would not have been able to dream of starting, never mind finishing a PhD! Thank you for being you and for being with me through all the ups and downs throughout these past few years.

Joe and Ana, thank you for being such great friends and PhD seniors to me! You made me feel so welcome in G.16, and were the catalyst for making friends with so many others. I am lucky that our friendship has gone past the PhD and will continue through into the far future. Mel, you have been another great friend I have made along the PhD journey, I am so pleased that James brought us together and we can keep his legacy going through our work!

To all my other colleagues in G.16, in the child language research group, and in the department generally, meeting you has been such an enriching experience, and you have inspired me in so many ways with your work and perspectives! It has been fantastic to know and work with you all.

Mum and dad, thank you for believing in me and letting me know how proud you are of me! You have helped me get through some of the hard times more than you know.

To all my friends, thank you for your enduring support over the past few years. You have all been so encouraging, and given me a lot of fun and laughter when I have needed it most.

I would finally like to thank the Economic and Social Research Council for their funding and support. They have allowed my, and many other people's research to happen which otherwise wouldn't have.

Thank you everyone, this thesis is dedicated to you.

Table of contents

List of tables

List of figures

Chapter 1. Background and Thesis Introduction: The influence of child and social factors on the efficacy of language interventions and their role as moderators of the effect of language on school readiness

This chapter will demonstrate the need for the current thesis by outlining the background to the topic issues. It will also summarise the core research question and introduce the research phases conducted to address it.

1.1 Thesis background

1.1.1. The developmental impact of preschool language and school readiness

Preschool oral language

Oral language is an important area of development for young children. From around 2 to 3 years old, children begin rapidly increasing their vocabulary, and start to use this to form short sentences, communicate needs and ideas, and converse with others (Ambridge & Lieven, 2011; Law, 2015; Law et al., 2017). Early oral language in turn enables a capacity to form relationships, and enables the child to interact with and learn from their environment (Ambridge & Lieven, 2011). This then results in developing more advanced language, socio-emotional and cognitive skills (Ambridge & Lieven, 2011; Hoff, 2013). As such, preschool language lays the foundation for future development and communication.

In contrast, there are serious consequences when children demonstrate preschool language difficulties. Longitudinal research has shown that children entering school with language difficulties are likely to have persisting problems with language development (Klem et al., 2016; McKean et al., 2017); and are at a significantly increased risk of poor outcomes in mental health, education and employment (Bishop, 2009; Bishop et al., 2016; Botting et al., 2016; Curtis et al., 2018, 2019; Feeney et al., 2012; Johnson et al., 2010; Klem et al., 2016; Law, 2015; van den Bedem et al., 2018).

Early difficulties with language are likely due to a cumulation of biological and environmental risk factors. For example, Eadie et al. (2022) found that when assessing the cumulative effects of early child (e.g., non-verbal cognition), and environmental (e.g., socio-economic disadvantage, home-learning environment) risk factors, the risk of poor language increased significantly the more risk factors they experienced. These consequences affect a significant minority of children. Norbury et al. (2017) reported that 7.6% of 4 to 5-year-old children without identified disabilities had difficulties with language development in a community sample. From these data they suggest that, on average, two children in every UK classroom have significant language difficulties. As such, a large minority of children are at risk of having poorer long-term outcomes due to low language abilities.

School readiness

Alongside preschool oral language, school readiness is a key indicator of development. School readiness is usually conceptualised as a multi-component construct that can be defined as being equipped with physical, cognitive, linguistic, and socio-emotional skills to learn and engage with school activities (Camacho et al., 2019; Duncan et al., 2007; Kokkalia et al., 2019; Law, 2015; Pan et al., 2019). While school readiness is used widely as an assessment of educational outcomes in research and for governments, it is complicated and controversial as a concept. This is because there is disagreement on what skills should be included, the theoretical basis for why certain skills are included under school readiness together, or how to appropriately assess school readiness skills (Kay, 2018; Snow, 2006). Further, some researchers are adverse to school readiness measures, as they argue that they generally 1) are not a realistic reflection of pedagogical practice; 2) not all children will fit neatly into devised benchmarks of success; and 3) place onus on the child to be ready for the curriculum and educational system, rather on the school being ready to teach the child (Kay, 2018; Pretti-Frontczak, 2014; Roberts-Holmes, 2019). As such, the definition of school readiness as defined in this thesis acknowledges these controversies, and describes school readiness more in terms as a measure of developmental benchmarking at a specific and important transitional point in a child's educational journey.

Research examining measures of school readiness have found children more ready for school are more motivated at school, and develop resilience towards new environments (e.g., classrooms), people (i.e., teachers and peers) and contexts (e.g., structured learning) (Bustamante et al., 2017; Law, 2015). Large-scale research exploring the effects of being ‘school ready’ consistently demonstrate that it predicts growth in educational outcomes for maths, literacy, physical, and socioemotional development throughout primary education (Davies et al., 2016; Pan et al., 2019; Ricciardi et al., 2021). In turn, the degree to which children differ in the constructs underpinning school readiness relates to better later adolescent and adulthood life outcomes in education, physical and psychological health, and criminal activity (Jones et al., 2015; Law, 2015; Rahman et al., 2018; Reynolds et al., 2011; Sadler et al., 2015). More specifically, poor achievement at preschool age has been linked to persistent educational, cognitive and socioemotional gaps with peers throughout school and beyond (Joshi et al., 2016; McCoy et al., 2017).

In the UK, children enter formal schooling at around age 5. Before this period is when the UK government assess and benchmark children's attainment and development to consider how ready children are to transition to Year 1. They do this via the early years foundation stage profile to assess school readiness at this transitional period (Department for education, 2020). The early years foundation stage profile is measured by teachers who are asked to score and qualitatively detail the extent a child has progressed or exhibited behaviours related to socio-emotional development, language and communication, academic achievement (e.g., literacy, maths), creative development, and physical development throughout the school year (Department for education, 2020). Generally, the government consider children are likely to be ready for school if they achieve at least average scores (set by the government) in socio-emotional, language and academic development (Department for education, 2020), and an overall average total score made up of these areas and the other areas of development measured.

With this benchmark in mind, many children in the UK are not school ready before year 1. In 2020, a UK government report (Nicholls et al., 2020) obtained data from both School Readiness and Teacher Track surveys, which asked teachers if children were ‘school ready’ or ‘not school ready’ based on government expectations. They found that around 12 pupils per class were considered to not meet the benchmarks of being school ready. Although COVID-19 may have contributed to this rate, the same report indicated 35% (around 9 students) did not meet the government benchmarks of school readiness in 2019 using the same survey. Additionally, government reports in the mid-2010s also suggested 42% of children did not meet the government benchmarks of school readiness (Office for Standards in Education, 2014). Together, these findings suggest over a third of UK children attending Year 1 are less likely to be ready for school, and this has been an issue for a number of years. This is of great concern, because this means a substantial number of children will be at risk of persistent issues and poorer outcomes throughout their lives. Consequences of poor educational outcomes have been found to create large costs to economic, health, and social systems (Davies et al., 2018; Joshi et al., 2016; Organisation for Economic Co-operation and Development, 2010). Therefore, should these rates of school readiness continue, both individuals and societies will be considerably negatively impacted.

The link between oral language and school readiness

Preschool oral language is a core component of school readiness, and is commonly examined in assessments of school readiness (Daily et al., 2010; Department for Education, 2020; Russo et al., 2019; Snow, 2006). This is because many school activities require adequate language to engage with tasks and instructions, and to understand specialised subject terminology at school (Collett, 2017; Schleppegrell, 2012). Oral language is also linked to, and impacts components which are often conceptualised as being part of school readiness. For example, good oral language underpins and predicts better performance in literacy, maths, and later language (Chow & Ekholm, 2019; Fuchs et al., 2018; Lonigan & Milburn, 2017; Trakulphadetkrai et al., 2020), and better behaviour and socio-emotional developmental skills like emotion regulation, social skills and behavioural problems (Bretherton et al., 2014; Chow et al., 2018; Levickis et al., 2018; Yew & O’Kearney, 2013). Good oral language is also shown to strongly relate to cognitive skills like processing speed and attention (Snijders et al., 2020; Willinger et al., 2019). Thus, both preschool oral language and school readiness are closely related, with early oral language being a subcomponent of school readiness, and a key factor influencing other domains which make up the 'school readiness' construct.

The association between child and social factors, and oral language and school readiness outcomes

As previously outlined, early difficulties with language are likely due to a culmination of biological and environmental risk factors (Eadie et al., 2022). Additionally, such child (individual attributes related to the child's development) and social (family and community-related experiences which develop the child through their environment) factors have demonstrated an association with oral language growth and children's readiness for school. More specifically, children's developmental vulnerabilities and social disadvantages have been shown to relate to poorer school readiness and language outcomes. For example, children with language difficulties and low performance in skills which comprise school readiness are also likely to be male, have socio-emotional difficulties, have a higher rate of health difficulties and developmental disorders (e.g., ADHD, speech sound disorders), live in poverty and deprived areas, and have parents with lower educational qualifications (Betancourt et al., 2015; Duncan et al., 2015; Duncan et al., 2007; Flouri et al., 2020; Hosokawa & Katsura, 2018; Levickis et al., 2018; Neuman et al., 2018; Paul, 2020; Vugteveen et al., 2021). As such, child and social factors may be important to consider because of their relation to attainment in both preschool oral language and school readiness outcomes.

1.1.2. Are current language interventions the answer?

Interventions targeting oral language and school readiness outcomes

The previous evidence for the consequences of poor language and school readiness has indicated that it is essential to intervene to promote these in order to improve life and societal outcomes. Commonly, preschool oral language and school readiness are addressed via psychosocially-based interventions. These consist of socially, psychologically, or cognitively based components aiming to improve oral language and/or skills which are often conceptualised as being part of school readiness (Enderby et al., 2013; Law et al., 2012; Ursache et al., 2012; Welsh et al., 2014). As such, clinicians and researchers have advocated for early language interventions, or school readiness interventions which also target language as a method to support school readiness (Leech et al., 2018; Perrin et al., 2020). Overall, such interventions have proved successful in improving a variety of different language abilities and/or social, behavioural and academic skills.

For example, Law et al. (2018) found in their systematic review and meta-analysis that parent-based book reading interventions promoted gains in pre-reading language skills, and expressive and receptive language. Furthermore, in a study examining an intervention curriculum targeting vocabulary, phonological awareness, print knowledge, and mathematics not only improved these outcomes, but also predicted better cognitive and socioemotional outcomes (Lonigan et al., 2015). In addition, a parent-implemented intervention targeting preschool phoneme awareness, vocabulary, narrative skills, and maths found improvements in children's language comprehension, vocabulary, academic skills (maths, literacy) and educational engagement (Noble et al., 2012). Moreover, two studies (Nix et al., 2013; Welsh et al., 2020) examining the effects of Head Start's REDI interactive reading program found targeting socio-emotional, literacy and oral language promoted gains in these areas. In addition, oral language and narrative comprehension gains were also seen to boost social-emotional skills and literacy for pre- and elementary-school aged children. Finally, in their review of oral language interventions, the Education Endowment Foundation (EEF, 2019) found targeting language has a high impact on improving academic attainment and reducing behavioural problems in children. Therefore, not only does promoting oral language create gains in oral language, but these improvements in oral language also promote gains in other school readiness skills.

Considering the implicit assumptions underpinning language interventions

But while intervention research demonstrates that preschool language interventions can benefit oral language and school readiness, their implementation and examination of effects tends to be motivated by set of implicit assumptions. Specifically for efficacy, they assume 1) children benefit equally from language interventions, and 2) children will benefit equally in school readiness from gains in language ability. This reasoning is evident when exploring how the efficacy of interventions are reported in research or evidence repositories. Specifically, the main focus when exploring effects tends to be assessing generally whether an intervention group's effect size is significantly different to control groups, the strength of the effect size, and/or the rate of growth since the beginning of the intervention began (e.g., as seen for the 'What works' repository from ICAN, 2021). Therefore, this means that groups are considered in a more aggregate sense, and there is little emphasis or exploration on which populations may be benefitting from language intervention and why. However, as noted in research outlined above, it clear that child and social factors relate to oral language development and school readiness achievement. Due to this, there may be numerous possible pathways or mechanisms through which oral language, school readiness, child and social factors associate with one-another. How they do so may undermine these implicit assumptions made when assessing and implementing language intervention. This is evident by the small number of intervention studies and systematic reviews with meta-analyses within the past 15 years, which have begun to examine the effects of child and social factors on intervention response, and are summarised here.

Roberts and Kaiser (2011) completed a meta-analysis on the impact of parent-implemented interventions for children (aged 1.5 to 5 years) with language impairment. They compared intervention effects between children with and without intellectual disability, and those with intellectual disabilities had smaller intervention effects on expressive vocabulary, but no other outcomes examined differed (overall language, expressive language, receptive language, rate of communication). When conducting a systematic review of 67 vocabulary intervention studies, Marulis and Neuman (2010) conducted moderation analyses comparing with and without 'at risk status' (at least 50% of the participant sample was within one risk category: 1) low SES level defined as at or below the national poverty level, parental education of high school graduation or less, qualification for free or reduced-price lunch; 2) second language status; 3) low academic achievement assessed by teacher reported or standardised school assessment; and 4) having an individualised education program or Title 1 placement) for vocabulary intervention gains. They also completed a comparison between children with different SES statuses (as described above = low) with middle to high SES children for vocabulary intervention gains. In both cases, while children with 'at risk' status and lower SES gained less in vocabulary outcomes, they were not significantly different to the not 'at risk' and middle- or high-SES children. Marulis & Neuman (2013) also completed a second systematic review and meta-analysis utilising 51 vocabulary intervention studies, and examined whether gains were moderated by the type of 'risk factor' (which could be defined by 1) marginalised ethnicity; 2) English Language Learners; 3) language difficulty; 4) low academic achievement; 5) low SES; 6) rural versus urban versus suburban). Only SES was a significant factor, with lower gains being found for children from low SES backgrounds. They also found a cumulative impact for low SES, where gains were reduced further with the addition of the other risks examined. A further systematic review and meta-analysis examining the impact of parent-child book reading interventions of language development and school readiness, (Law et al., 2017) found that although child-parent reading interventions generally provided positive outcomes for language, effects were stronger for more socially disadvantaged groups. Dowdall et al. (2020) also examined shared book reading interventions, and found age did not moderate expressive or receptive language outcomes. Finally, a randomised controlled trial by Boyle et al. (2007) compared the gains in expressive and receptive language for interventions with different implementers (SLT, SLT assistant) and modes (1:1, small group) for older primary school children (aged between 6-11 years) with language difficulties. They examined the moderating effect of language profile (expressive, receptive or mixed language difficulties), gender assigned at birth, and NVIQ. Non-verbal IQ did not moderate intervention response. Gender assigned at birth and language profile (expressive versus mixed) did moderate gains in receptive vocabulary, with females gaining more than males, and children with expressive language difficulties gaining more than those with mixed language difficulties (expressive language could not be modelled satisfactorily for these analyses).

However, it could be argued that carefully considering intervention ingredients may help ameliorate child and social level differences outcomes equitably for children. Research has clearly demonstrated intervention efficacy is related to intervention ingredients like dosage and more direct implementation (e.g., Frizelle et al., 2021a, 2021b; Tosh et al., 2017). As such, if intervention is applied optimally, it may benefit children regardless of their social disadvantage and developmental vulnerabilities. However, equity cannot be achieved solely by adjusting intervention components. This is because language and school readiness interventions are complex interventions, defined by having a high number of complicated and contextual interactions between components and experimental groups, plus each individual within them (Skivington et al., 2021). Skivington et al. state these differences are key in understanding different efficacy levels seen in interventions, and so will be important to consider to produce the most benefit for the populations they serve. In other words, while intervention ingredients can partially contribute to equitable efficacy, only considering these do not fully consider the complex nature of different populations. An example of this is demonstrated by case study research by Storkel et al. (2017) who found that children with more severe phonological awareness, vocabulary and non-word repetition difficulties were less likely to make gains in vocabulary from their interactive book reading intervention despite being provided an optimal level of intervention dosage. As such, individual contexts (i.e., social disadvantage and developmental vulnerabilities) could be an important source of unequal intervention response that needs to be addressed.

The second assumption, that having better language means all children benefit equally in school readiness outcomes may also be flawed. In addition to longitudinal research of child and social factors individually predicting components of school readiness, Prior et al. (2011) and Hammer et al. (2017) both examined children in the Early Language in Victoria Study cohort. They found socio-economic status and language could both individually contribute to school readiness growth at the same time. Furthermore, in a study analysing the British Cohort Study data, Feinstein (2003) found children from lower socio-economic backgrounds with higher scores on tests of language, cognitive and socio-emotional development as toddlers demonstrated less growth in these skills in later childhood compared to their more affluent peers. This may indicate that not only do risk factors have an effect on school readiness independent of language, but they may also affect the ability of children to capitalise on initial language advantages.

Why these findings may be of particular concern is because they could mean children get a cumulative ‘triple threat’ of disadvantage from developmental vulnerabilities and social disadvantage. That is, oral language and school readiness may be affected by 1) direct effects of social disadvantage and developmental vulnerabilities, 2) poorer response to language interventions and 3) less benefit accrued for school readiness from language gains. In other words, children with developmental vulnerabilities and social disadvantages could be receiving a cumulative disadvantage towards their language and school readiness development, gains from intervention, and gains in school readiness even if they benefit from interventions. If this does occur, then employing current interventions without considering how to tackle these levels of disadvantage will only compound difficulties that children have.

Research gaps

While the current evidence is indicative of child and social factors affecting response to language interventions, it is limited and subject to a number of research gaps. First, the pool of child and social factors examined were limited in studies. The majority of these reviews/studies focus mostly on social disadvantage factors (predominantly socio-economic status), while each study/review generally focused on a single or small number of child factors. Although social factors are clearly important, much more work on child-level factors is also needed. In addition, findings relating to most factors were from small samples, and/or a small number of studies. For the reviews and single intervention studies, effects of child and social factors on intervention efficacy were generally not the focus of the studies, but instead on the effectiveness of a particular intervention type (e.g., parent-child reading, vocabulary-based interventions). In their review, (Law et al., 2017) recommended that more research needed to be completed for different intervention types, factors, and different populations. Furthermore, some effects were found for older children (Boyle et al., 2007), while results could be different for preschool-aged children. As such, a more comprehensive and focused examination of the effect of child and social factors on preschool intervention response is needed.

In addition, no research to my knowledge examines the potential moderating effect of child and social factors on the relationship between oral language and school readiness. Research is currently based on separate associations between 1) factors and oral language, 2) oral language and school readiness, and 3) factors and school readiness. As such, more longitudinal research is needed that utilise predictive interactive models to understand how changes in child and social factors affects benefits made in school readiness from gains in oral language.

1.2. Research questions, methods chosen and thesis structure

There is a clear need to understand whether child and social characteristics 1) affect language intervention response, and 2) affects their school readiness outcomes from gains in language ability. The overarching research question is:

To what extent do child and social factors moderate the efficacy of language interventions, and what is their role as moderators of the effect of language on school readiness?

The approach to enquiry is positivist, and specifically based on biostatistical, epidemiological, psychological and health sciences fields. The thesis is split into two phases, using empirical methods to answer two research questions:

Phase 1: Do children benefit equally from interventions, or are gains affected by child and social factors?

This question is addressed in chapter 2 through a systematic review. This review synthesised data from language intervention studies treating preschool children with language difficulties. Results are presented of studies using analysis methods I have characterised as ‘third variable’ analyses. These were defined as analyses including at least one additional (child or social) variable(s) in the analysis to the predictor (first variable) and outcome (second variable) which may be driving additional changes in the outcome (detailed further in chapter 2). Types of ‘third variable’ analyses included for consideration were subgroups, correlation, covariates, moderation or mediation.

Phase 2: Do children benefit equally in their school readiness outcomes from gains in language ability, or are these benefits moderated by child and social factors?

This question is addressed in chapter 3 through analysis of longitudinal data from two waves (age 3 and 5) in the nationally representative Millennium Cohort Study. The cohort included data for oral language (expressive vocabulary), child- and social factors, and school readiness (Foundation Stage Profile) for preschool-aged children with a spectrum of language abilities.

Finally, in chapter 4, the findings from both phases are brought together and discussed to identify key implications, recommendations and future directions for research, policy and practice.

Chapter 2. The impact of child and social factors on the efficacy of language interventions: A systematic review and narrative synthesis

2.1. Background and research aim

Chapter one outlined the research issues for the thesis, and so a brief overview specific to this phase is highlighted here. This section will also outline the choice to conduct a systematic review for this phase. Furthermore, studies included in the review were analysing child and social factors in a number of different ways. As such, this section outlines and describes these different types of analyses, named ‘third variable’ analyses in this chapter. Finally, the research aim is reported.

2.1.1. The potential impact of child and social factors on language intervention response

Language difficulties are associated with poorer outcomes for children’s long- and short- term educational and life outcomes (Bishop, 2009; Bishop et al., 2016; Botting et al., 2016; Curtis et al., 2018, 2019; Feeney et al., 2012; Johnson et al., 2010; Klem et al., 2016; Law, 2015; McKean et al., 2017; van den Bedem et al., 2018). Due to this, utilising interventions targeting language to prevent such issues is considered essential, and have been widely used and successful (EEF, 2019; Law et al., 2018; Lonigan et al., 2015; Nix et al., 2013; Noble et al., 2012; Welsh et al., 2020). However, language intervention tends to be motivated by an implicit assumption that all children will benefit equally from language interventions. This may be flawed as research indicates that developmental vulnerabilities and social disadvantage not only place children at risk of poor oral language development, but may also impact their intervention response (Boyle et al., 2007; Dowdall et al., 2020; Law et al., 2017; Marulis & Neuman, 2010, 2013; Roberts & Kaiser, 2011; Storkel et al., 2017).

However, the amount of evidence for selected child and social factors utilised in moderation analyses for review studies was limited. This was because the pool of child and social factors were limited in studies, and findings for most factors were from small samples, and/or a small number of studies. The effects of child and social factors on intervention efficacy were also generally not the focus of the studies. As such, the current analysis aims to conduct a comprehensive approach, and focus specifically on the effects of child and social factors on language intervention response.

The inclusion of child and social factors was based on prior literature examining associations with language development, or based on previous studies examining factors' impact on intervention efficacy (highlighted above and detailed in chapter 1). Factors examined were initial language ability, language profile, non-verbal IQ, co-occurring disorders, age, gender assigned at birth, and socioeconomic status. These are reported and hypotheses for each are presented in section 2.2.6.

2.1.2. Choosing a systematic review

This phase assessed whether children benefited equally from interventions, or if gains were affected by child and social factors. In order examine this, a systematic review and narrative synthesis was completed. This method was chosen as it could provide a comprehensive overview of the current preschool language intervention literature (Moher et al., 2015). This was important to help explain what may be creating differential outcomes in language interventions, and to provide recommendations for what factors need to be identified and addressed in future interventions. It also helped inform present research gaps and requirements for future research. Chapter 1 highlighted that there are likely to be a number of factors which have not been examined, or where data has not been drawn together to establish how these factors relate to language intervention outcomes. In support of the main research aim, methodological concerns in studies were explored to determine that the robustness and generalisability of conclusions drawn from studies and collated evidence. The current systematic review will be reported in line with the latest PRISMA guidelines for systematic reviews (Page et al., 2021).

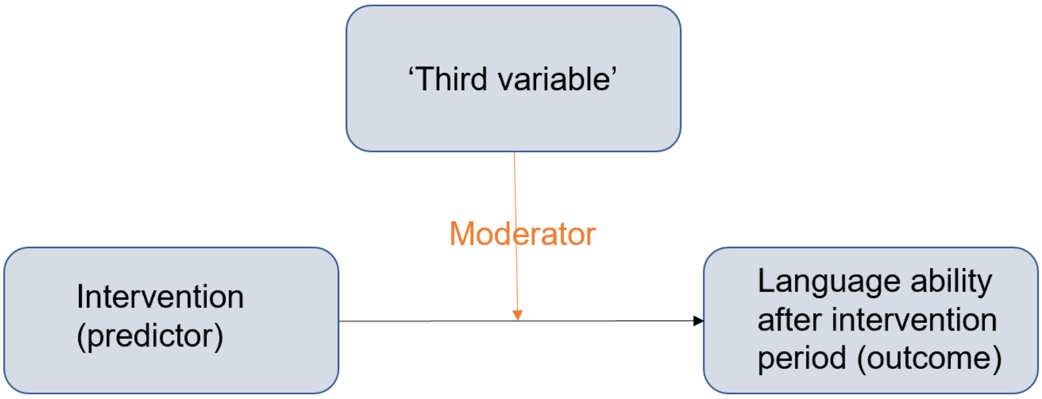
2.1.3. 'Third variable' analyses

As seen in chapter 1, the amount of evidence for selected child and social factors utilised in moderation analyses for review studies was limited. In addition, it has been noted that many language intervention studies have narrow inclusion and exclusion criteria, removing children with broader difficulties (Law & Stringer, 2014). As such, it was considered that this may also be the case for individual intervention studies. Thus, an inclusive approach was employed to analyses, as this would allow for a more informative synthesis of available data on how factors affect intervention response. Therefore, relevant analyses other than moderation were also included if available. Analyses included were those that could demonstrate a relationship to language intervention outcomes, and are dubbed as ‘third variable’ analyses in this thesis.

‘Third variables’ are defined as additional variables to the main predictor (first variable) and outcome (second variable). For the purpose of this review, the main predictor is participation in the intervention (or not), and the outcome is the oral language ability measured after the intervention. The ‘third variable(s)’ are the child or social factors. The choice of the term ‘third’ does not just mean there is necessarily only one additional variable (there can be multiple), but indicates the presence of a third type of variable (e.g., predictor, outcome, moderator). ‘Third variable’ analysis is therefore an umbrella term used to refer to potentially different ways child and social factors are entered into the analysis alongside the main predictor and outcome variable. For example, a ‘third variable’ analysis that could be included is how socio-economic status moderates the relationship between language intervention and children’s language outcomes. However, how factors are entered into the analysis differs, meaning findings produced from them have different implications. To understand each ‘third variable’ analyses included here, it is important to recognise exactly how they are entered alongside the main predictor and outcome variables, and what they can do. The ‘third variable’ analyses selected here are based on those commonly utilised in social science (and language development and disorder) research. Path models for each ‘third variable’ analyses will be provided in figures 2.1 to 2.5, and a combined path model is provided in figure 2.6 to give a visual overview of how they work compared to each other.

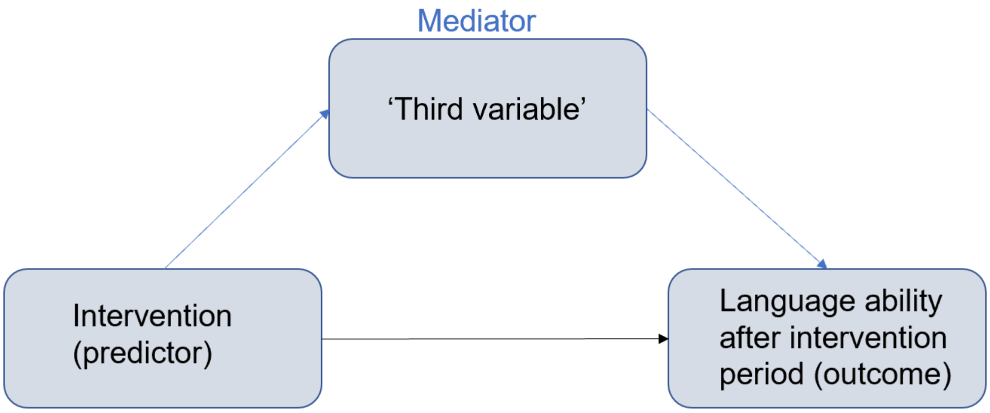
Moderators were first described within the social sciences by Baron and Kenny (1986) as variables which interact with the predictor to produce different effects in the outcome. Specifically, the different levels of the moderator variable changes the direction and strength of the effect of the predictor on the outcome (Baron & Kenny, 1986; Bhandari, 2021). Simply put, the improvement from the intervention differs according to differences in the level of a factor. For example, males may respond differently to an intervention than females, which creates a differing intervention efficacy for each group.

Figure 2.1. Path diagram illustrating moderator analysis



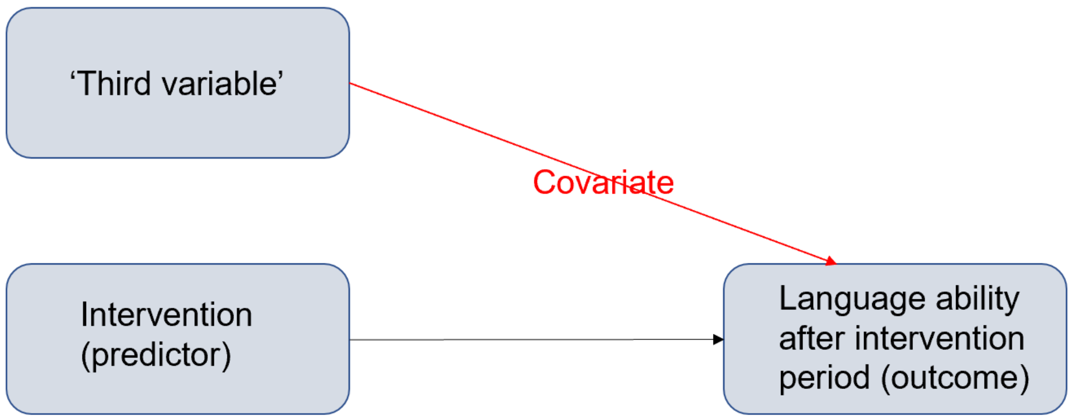
Mediation is an analysis where a 'third variable' intervenes between the predictor and outcome, and is the true explanation of the relationship (Bhandari, 2021). That is to say, the predictor creates changes in the child/social factor, which then influences the outcome. Therefore, mediators explain the level of change from the predictor to the outcome. For example, the intervention (predictor) may change the cognitive processing of children ('third variable'), which then creates a different intervention response (outcome).

Figure 2.2. Path diagram illustrating mediator analysis



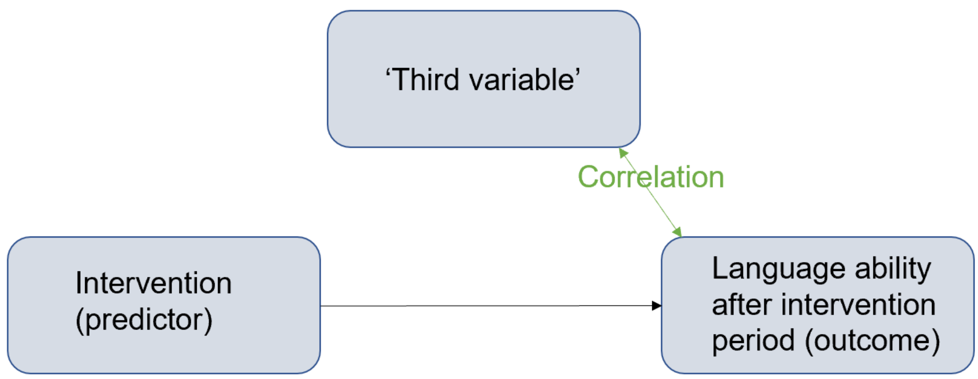
In addition to moderation and mediation, a further three 'third variable' analyses were considered. However, their relationships between predictor and outcome is less clear and more limited. Covariates (described by Kim, 2018) can demonstrate how a factor individually predicts the outcome, and can outline how much (via %) of the outcome is explained by them. Depending on the type of analysis (e.g., in a regression model), it can also indicate the direction of the relationship with the outcome. However, they do not interact with the predictor variable, which means they are unable to explain how different levels in the factor affect the outcome. For example, in addition to the intervention, socio-economic status accounts for some differences in intervention efficacy. However, it is unclear which level(s) of socioeconomic status this applies to, and how these groups individually respond to intervention.

Figure 2.3. Path diagram illustrating covariate analyses



Correlations (described by Kim, 2018; and Schober & Schwarte, 2018) can demonstrate the simple relationship between a factor and outcome. It can also indicate the direction of the relationship with a variable. For example, cognitive ability relates positively with intervention outcomes, which means a higher cognitive score relates to a higher intervention outcome score. However, it would be impossible to ascertain if there are additional variable(s) influencing this relationship. Moreover, it only explains the extent of how variables are related, and so cannot explain how different levels in the factor relate to the outcome. However, they may still be useful in finding relationships for further exploration in future studies.

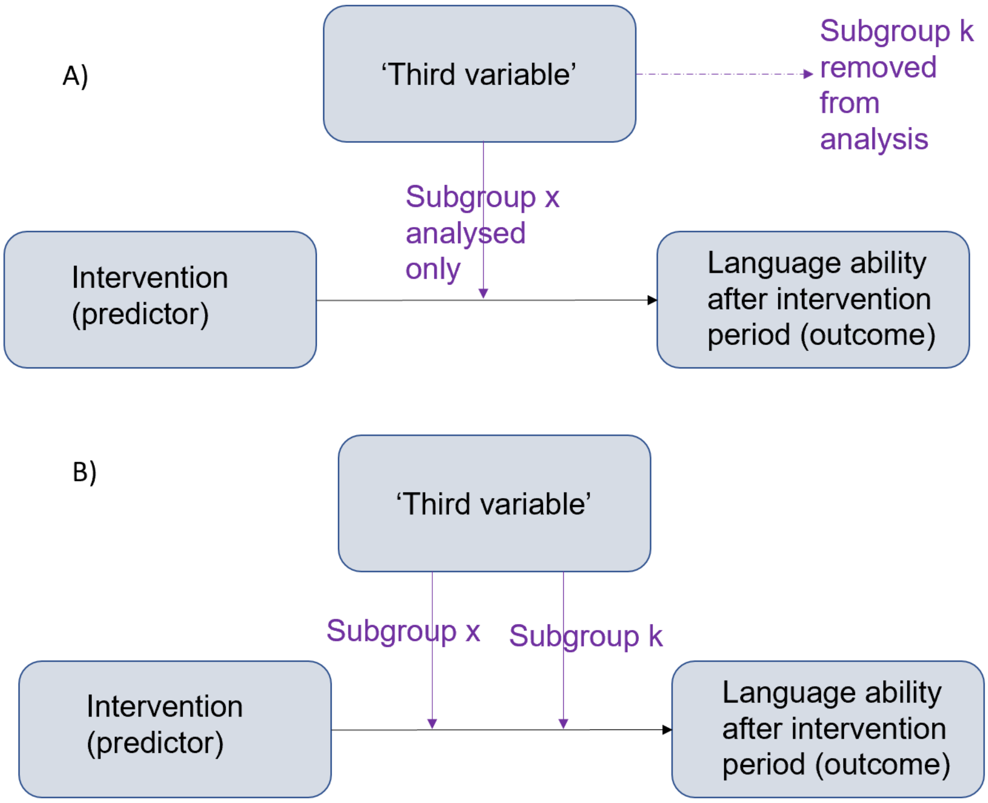
Figure 2.4. Path diagram illustrating correlation analysis



Note. A correlational relationship could also occur between the predictor and factor. However, these relationships are not included in the current phase as they cannot inform differential outcomes.

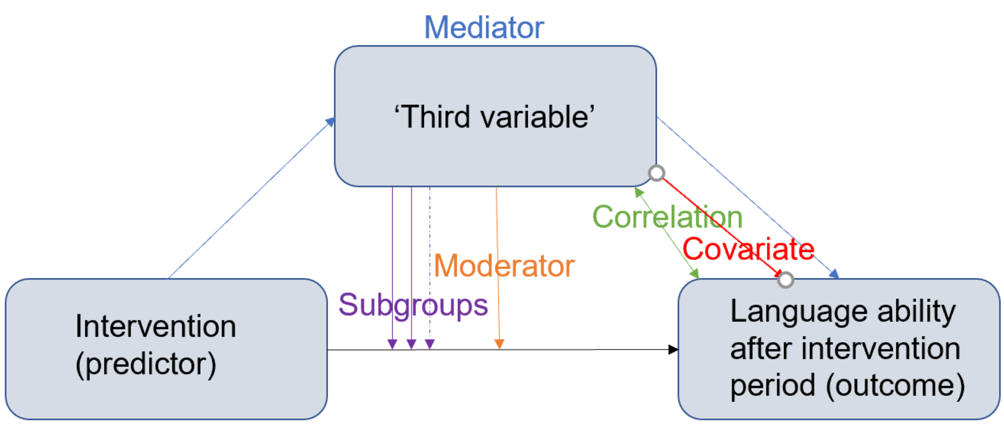
Finally, findings dubbed under subgroup analyses seem to be very similar to moderation analyses (different groups potentially showing differential intervention outcomes), but are distinct from it and draw less robust conclusions. Subgroup analyses here could fall under two different types. The analysis procedure involves either 1) removing subgroups from the main sample, or 2) conducting separate analyses for each subgroup. These are then compared with the initial intervention efficacy results (and/or between subgroups). An example of the two subgroup analyses are 1) removing children with ADHD from the sample in an analysis and then comparing the intervention effect size to the one with all children in, or 2) analysing monolingual children in one ANOVA, analysing multilingual children in another ANOVA, and then comparing the results of each ANOVA with each other (and also an ANOVA with both groups in together). Analyses splitting groups like this typically occur when variables outside of the predictor and outcome are found to be significant, or are identified as potentially influential variables in the analyses *post hoc* (Frey, 2018). This means that these analyses should be treated with more caution than conclusions from moderation analyses. One reason for this is because the factor had not been fully considered before analysis, meaning they are not based on established theory (Frey, 2018). Further, a methodological issue with splitting up samples into smaller sizes or removing participants is that this can inflate effect sizes (Slavin & Smith, 2009). Together, these issues means that the estimated effect sizes calculated from this type of analysis may be subject to bias. However, they may still be useful in finding tentative relationships to explore in future studies.

Figure 2.5. Path diagram illustrating subgroup analyses



Note. A) is a removal of one subgroup from the analysis to determine how it affected the intervention efficacy, and B) separate analyses for each subgroup of the factor.

Figure 2.6. Path model of the predictor-outcome relationship with all 'third variable' analyses



2.1.4. Research aim

To determine if children benefit equally from interventions, or if gains are affected by child and social factors

To do this, the review synthesised data in two phases to examine:

1) What participant factors are described in intervention studies for preschool language, and have been included in ‘third variable’ analyses, and why?;
2) What conclusions can be drawn regarding the impact of the identified child and social factors on preschool language intervention response?

2.2. Methods

2.2.1. Ethical approval

Ethical approval for the systematic review was granted from Newcastle University’s Humanities and Social Sciences Ethics Committee in December 2017.

2.2.2. Eligibility criteria

Eligibility criteria for study inclusion were designed with reference to the PICO reporting framework where criteria for participants, interventions, comparators and outcomes were considered. In addition, considerations for criteria for study designs, publication types, date range and language, and ‘third variable’ analyses are also reported. It should be noted comparator (i.e., control group) information is described within the participants and study designs sections where relevant. Criteria for each aspect are outlined below together with their rationale. As well as the above criteria, an ‘unsure whether study should be excluded’ option was included so that papers where it was unclear whether to include papers could be reserved for discussion between JT and the supervisory team (Prof. Cristina McKean (CMK) and Prof. James Law (JL)) later (see the selection process section, 2.2.4 for more information). An overview of the eligibility criterion is given in table 2.1.

Participants

Studies were chosen if children had, or were at risk of, language difficulty. This was defined in two ways. First, children could meet the diagnostic criteria of Developmental Language Disorder (DLD). All criteria for, and specific statements about DLD are outlined in Bishop et al. (2017). But briefly, DLD is defined as a persistent and significant set of difficulties (spoken and/or understanding) in oral language in all the languages a child speaks. DLD would not be diagnosed if language difficulties are present from/ alongside differentiating conditions (such as autism or intellectual disability), as the language difficulty is likely part of a set of complex impairments. DLD would also not be diagnosed if a child had isolated speech-sound disorders.

Additionally, children could also be at risk for DLD. Although risk for DLD varies based on a child's age, preschool aged children may be considered 'at risk' if the child is demonstrating impaired language (usually in multiple areas) in testing, already receiving intervention services, or have specific family factors such as a family history of speech and language difficulties and belonging to families from low SES backgrounds (Bishop et al., 2017). Therefore, children were also included if they did not have a formal diagnosis but poor performance on language measures, and/or inclusion in certain social groups which are at risk for poor language. Regardless of diagnosis or 'at risk' status, a language difficulty was defined by a score of -1SD or below. This decision is further detailed in section 2.2.6 detailing initial language ability. Studies were also included only if children were aged on average between 3;0 and 7;0 years old. Before age 3, it is very difficult to determine if a child has or is at risk of DLD (Bishop et al., 2017). The upper age limit of 7 years encompasses the entry to school in the UK an in educational systems of other countries which have kindergarten provision.

Children in the intervention and comparator groups both had the same eligibility criteria. For example, controls could not be typically developing or matched for language (e.g., younger children scoring the same on measures). This was because comparing outcomes of groups similar to one-another is essential to clearly understanding intervention efficacy and the impact of child and social factors. In other words, if groups performed differently on outcomes, it could not be determined if this is because of the different developmental profiles of the samples, or due to the intervention and/or child and social factors.

Interventions

To be included in the review, studies had to examine interventions aiming to promote language development for at least one oral language skill. The intervention was required to be 1) non-pharmacological and/or surgical; 2) socially, psychologically, cognitively or educationally-based (i.e., psychosocial) interventions; and 3) involve an implementer who aims to improve a skill or behaviour with the patient. However, an inclusive approach was employed for service delivery issues, allowing any implementer of the intervention (e.g., parent, teacher, clinician etc.), any dosage amount, and regardless of whether the intervention was implemented via face to face or by remote methods. The only restriction was that interventions must be overseen by a language specialist such as the experimenter or a speech and language practitioner. This was because language specialists are able to provide evidence-based support and training to implementers, as well as being able to track outcomes. As will be seen in section 2.2.6 (intervention and control information items) a string for reading interventions was added in the search strings. While not examining literacy outcomes, this type of intervention was included because reading tasks are commonly part of, or the main component in many oral language interventions (e.g. Mol & Bus, 2011), and so it was important to ensure these types of intervention were included as they can promote oral language outcomes.

Outcomes

The primary outcome was oral language. This included any measures of 1) expressive language; 2) receptive language; 3) vocabulary and word knowledge; 4) spoken language comprehension; 5) pragmatics; 6) grammar; 7) morphology; 8) narration; 9) phonological awareness/ knowledge; 10) general language (defined in the thesis as expressive and/or receptive language difficulties identified by omnibus language measures). ‘Speech’, defined here as abilities involving physical movements for talking and sound articulation, is not examined as an outcome, but as a potential child level factor detailed later in section 2.2.6.

Study design

To be included, studies had to employ either randomised controlled trials (RCT) or quasi-experimental study (QEs) design. RCTs involve random allocation to either an intervention or control group to study its effects (Gillam et al., 2008), and are a gold-standard for research quality in terms of their methodological procedure and reporting (Bothwell et al., 2016; Miller et al., 2020). For example, as participants are randomly allocated to groups, this reduces the possibility of selection bias, and so provides more confidence in the outcome. QEs also compare intervention and control groups but provide no randomisation, and generally have smaller samples and recruit participants from a smaller pool than RCTs (Miller et al., 2020). As such, RCTs are likely to be more generalisable compared to QEs. QEs are more susceptible to selection bias and so provide less confidence in outcomes (Thyer, 2012). However, there are only a small number of RCTs for child language and especially language intervention studies. Therefore, just including RCTs would have limited the number of available studies considerably. In contrast, QE designs are utilised commonly for language intervention, and so were included as they best reflect the current status of the field, as well as to increase the potential number of studies examined. Additionally, RCTs and QEs were likely to have sufficiently similar methodologies and reporting standards, and therefore had the highest potential for meta-analysis (Bärnighausen et al., 2017; Kabisch et al., 2011). QEs also have many advantages over RCTs. They can provide detailed contextual information of studies which is generally not present in RCTs, and are a popular and cheaper choice for conducting smaller scale intervention research (Gopalan et al., 2020; Miller et al., 2020). Therefore, they were chosen for their higher likelihood to complete 'third variable' analyses. Furthermore, Handley et al. (2018) state studies such as QEs are advantageous over RCTs in that they are better focused on small clinical subpopulations which are most likely to require treatment. They can also be used as a smaller scale assessment of treatment before being applied to a larger population. As such, their data provides a better balance of internal and external validity, and also complement RCT findings (Geldsetzer & Fawzi, 2017; Handley et al., 2018).

Research designs excluded were non-intervention research, single or multiple case studies and series, and studies where participants were their own comparator (e.g., pre-and-post tests, ‘before and after’ studies). Case study/series are an in-depth and detailed examination of participants involved, and provide large amounts of contextual information on research conditions (Lobo et al., 2017). Although the level of context could be useful for answering some of the aims of the review, Lobo et al. describe these studies as consisting of small numbers of participants, having no randomization or comparator, and reporting primarily qualitative data. This means that results are highly likely to be subject to bias and difficult to generalise. Studies were included if the comparator groups were no treatment, treatment as usual, or a delayed treatment. Alternative interventions as controls (i.e., non-inferiority trials in terms of content, not dosage) were also included. Studies where children were their own comparator were excluded because it is difficult to determine how effective the intervention is in general when it is compared against itself, and so suffers from issues of both internal and external validity (Knapp, 2016).

Publication types

Only fully reported studies were included. Non-empirical materials (e.g., editorials, correspondences, reviews, books, and book reviews), incomplete study information (e.g., protocols, conference abstracts or proceedings, research summaries, or only the abstract or study reference could be found), and undergraduate or masters’ dissertations were also excluded as sources. Although potentially informative, non-experimental materials are either more theory-based, non-empirical, contain studies which do not fall under the above inclusionary criteria or are a collection of studies. Undergraduate and masters theses were excluded because the level of scrutiny is inconsistent and unlikely to be at a level equivalent to peer review. However, PhD theses were included for consideration due to their work being deemed to a publishable standard within a peer reviewed journal.

Date range and language

Studies published between January 1st 2002- December 31st 2018 were included. Research within the last fifteen years was chosen to represent the most recent intervention and research practices and findings of the field. As the current project is a thesis, analysis work is to be completed solely by the author, JT (monolingual English speaker), only English language papers were included for analysis.

'Third variable' analyses

Papers were included if they conducted at least one of the 'third variable' analyses described in section 2.1.3.

Table 2.1. Final criteria for screening on title and abstract

Criteria	Specification
Include	1) Participants fall under the criteria of language difficulty (diagnosed or are at risk of DLD, Bishop et al., 2017); AND 2) Average age of participants is between 3;0-7;0 years old; AND 3) The study is a randomised controlled trial OR a quasi-experimental study for a language intervention; AND 4) Measures oral language outcomes (as specified in section 2.1.3) OR 5) Unsure whether study should be excluded (required comment from reviewer to inform discussion with the review team)

Criteria	Specification
Exclude (E1): Studies with the wrong groups	If participants do not fall under the DLD criteria (Posited in Bishop et al., 2017). Examples include: 1) Disorders considered a “differentiating condition” (e.g. Autism, Downs Syndrome, Intellectual Disability (NVIQ under 70 or specific diagnosis), other general learning difficulties, brain injury, acquired epileptic aphasia in childhood, neurodegenerative conditions, cerebral palsy or sensory-neural hearing loss); AND/OR 2) A participant only has a phonological difficulty or speech disorder (i.e. SSD, dyspraxia, stuttering).
Exclude (E2): Studies that are not interventions	1) A specific type of literature (Editorials, correspondence, reference or abstract available only, protocols, research summaries, books, book reviews); AND/OR 2) A type of research which does not explicitly examine a psychosocial intervention for oral language (Screening, prevalence, ‘disorder-explorative’).

Criteria	Specification
Exclude (E3): Studies that are interventions but of the wrong type	1) Studies using pharmacological and/or surgical interventions; OR 2) If the study does not examine any of the oral language outcomes (as specified in section 2.1.3).
Exclude (E4): Studies which included groups where the average age was either too low or too high	1) If the participants are on average age younger than 3;0 years OR older than 7;0 years.
Exclude (E5): Studies that are interventions but do not meet inclusion criteria on the grounds of the methods used	1) If the study is not a randomised controlled trial; OR 2) If the study is not a quasi-experimental study.

Criteria	Specification
Exclude (E6): Other reasons	1) If it is either an undergraduate or masters study; OR 2) If the study is published before 2002; OR 3) If the paper is not written in English;

2.2.3. Search strategy and information sources

Developing search strings

Search terms were utilised to collect studies and the search strings for each database are provided here. They were adapted from Law et al. (2017) and modified to fit the purposes of the current review. These modifications were to add strings for reading interventions, quasi-experimental studies, and reviews. The review string was included to acquire additional literature for the thesis, and to ensure the current review was not a duplicate of any previous ones. Another string was added for the dates focused on for this review (2002-2018). Strings relating to adolescent samples and drug therapies were removed, as these were not of interest for the current review.

Information sources

Studies were identified via the following sources. Bold text indicates the database/ source name, and is followed by the provider of the database/ papers and the specific catalogued resource used (if applicable):

1) **MEDLINE** (Via Ovid) [Ovid MEDLINE(R) Without revisions 1946 to November week 4 2018]
2) **Embase** (Via Ovid) [Embase 1974 to 2018 December 01]
3) **PsycINFO** (Via Ovid) [PsycINFO 1967 to November week 4 2018]
4) **ERIC** (Via EBSCO)
5) Scopus
6) **CENTRAL** (Via Cochrane)
7) **Web of Science** ['Web of Science Core Collection' = Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI]
8) **ETHOS** (Via the British Library) – Database used to find unpublished postgraduate (PhD) theses

9) **Hand-searching** reference lists of included studies.

10) Communication with leading academics in the field.

Sources 1-7 are databases which catalogue language disorder and language intervention research. Source 8, ETHOS, is a UK-based repository with over 500,000 theses, and so was important for sourcing potentially relevant PhD theses. Papers were accessed via databases in December 2018. Sources 9 and 10 were accessed at the end of the full-text screening phase. Experts were selected by both recommendations from the supervisory team (CMK and JL), and for authors whose work is prolific in the field. Emails were sent out to experts in June 2019 with information about the project as well as what type of papers were needed.

2.2.4. Selection process

Study selection via title and abstract and full-text screening was completed. These are outlined below together with their procedures. The main reviewer was myself (JT) with assistance from the supervisory team (CMK and JL). None of the reviewers were blinded to the name(s) of the author(s), institution(s) or publication source at any level of review.

Title and abstract screening

To ensure the screening process was valid and reliable, a pilot of 20 random papers were screened by both JT and JL and then discussed. Changes were made to the criteria due to issues such as being too cumbersome or overlapping with each other. This resulted in clarifications for a second stage of piloting. A further 20 papers were coded (i.e., assigned a decision based on criterion) to test their efficacy, and some final more minor changes were made and agreed on by the full review team (e.g., there was an ‘Include: Unsure’ criterion which was removed, and reviewers instead used the include code and made a comment on the paper if they were unsure). The level of agreement during the piloting stages were 75% for the first pilot (between JL and JT), and 95% between all researchers (JL, JT and CMK) at the second pilot stage, an acceptable level of agreement. Due to the level of agreement at the second pilot stage, the minor changes were added without further piloting. The final title and abstract screening criteria are presented in table 2.1.

After piloting, the title and abstract of papers from 4,543 papers were screened by the review team. Due to time and resource restraints, CMK and JL split the double screening for a proportion of the papers to ensure JT was adequately screening papers, and codes were still comprehensive enough. Over 15% (17.75%) of papers were double screened at title and abstract, with 564 of the database papers initially being double-screened by JT and CMK and all ETHOS theses being screened by JT and a combination of JL and CMK (206). Any disagreements were discussed and resolved. The agreement rate was above 95%, and so adequate for JT to continue title and abstract screening alone.

Full text screening

After screening the title and abstract of papers from all sources, full text screening was completed for 231 papers. Again, due to time and resource restraints, CMK and JT double screened for a proportion of the papers (24, around 10%), with disagreements discussed and resolved. Of the papers double screened, the agreement rate was above 95%. Agreement levels were again high, so it was agreed that JT could continue alone for the rest of full-text screening process with guidance from supervisors when required. For papers that JT was unsure should be included, these were shared with CMK and JL who helped confirm or exclude these papers through discussing the areas of uncertainty. This was done by both going through the paper and referring to the criteria, and if any uncertainty occurred (e.g., it was unclear whether children in the sample had differentiating conditions or not), an attempt was made to contact the author to clarify. If the author could not be contacted, then a decision was made based on information available. Once this was finished, the final papers were again looked over by JT, utilising the final criteria to determine if there were any further uncertainties. This did not occur.

2.2.5. Data collection and management

Title and abstract and full text screening were completed on EndNote© X9 and Microsoft Excel© software. If references were from databases, they were imported via .RIS files to EndNote. If references were from ETHOS, acquired by hand searching or by expert communications, these were created and screened in Excel. This is because creating individual references was significantly more time-consuming In EndNote, and so using Excel increased the efficiency for examining hundreds of papers. Screening in EndNote was completed by inserting extra fields into all Endnote reference templates (reviewer's initial, decision and comments if needed) which could then be filled in and compared using the smart groups feature. A similar set of information was recorded in Excel with additional columns for the information next to the reference. Data extraction and synthesis was carried out using Microsoft Word© and Microsoft Excel© because they are simple to set up data extraction codes in and navigate. Raw data extraction was completed via a data extraction form created by JT. To ensure the best quality data and all relevant information was acquired, CMK and JL reviewed data once JT had collected it.

2.2.6. Data items

The following items were extracted from the papers and followed PRISMA guidance for covering PICOs (Participants, Interventions, Comparators and Outcomes). Additionally, codes related to 'third variable' analyses, and child and social factors were added. Data was extracted from the included papers according to the requirements of the review research aim and objectives. Data extraction fell under five categories; 1) study design information (e.g., authors, study type, number of participants), 2) 'third variable' related sample information (i.e., how selected child and social factors in samples are being described, and what of this is presented in inclusionary criteria or as additional sample information), 3) intervention and control information (i.e., intervention and control group details and procedures), 4) outcome information (i.e., the oral language skills, measures and scores examined in each study) and 5) study findings (i.e., intervention-control comparison analyses, 'third variable' analyses).

Study design items

Study design items were chosen to provide a description of the included papers. Four study design items were extracted. These were study design (QE/RCT), country of study, participant numbers included at final analysis and whether individual or group data was provided. The study design information was required to support decisions regarding comparability of studies and therefore how possible meta-analysis was. For example, QEs and RCTs cannot be placed in the same meta-analysis, as they are too dissimilar. Knowing study type also helps with exploring the generalisability and robustness of findings (e.g., RCTs are less prone to bias). Country of study were also extracted. Language difficulty is prevalent across the world, and has been studied over multiple languages and cultures (Law et al., 2019a; Thordardottir, 2010). Although language difficulty is acknowledged globally, the funding of, and theoretical paradigms of causality and treatment vary across countries and cultures (Law et al., 2019a). Due to this, this information was utilised to consider potential publication bias (i.e., whether results fit within the contexts of specific countries or cultures). Country was based on where the study was completed rather than where it was published. No data was missing or unclear for design and country information, although some information was more implied in some studies than others (e.g., the reader could gauge a study is an RCT by how its design was described). Next, the final analysis numbers were extracted to determine study sizes (and therefore potential generalisability), and to help calculate effect sizes. If data about number of children included in the final analyses were missing, the total number of children reported would be used to calculate effect sizes. If there was an unclear number (i.e., a possible range), then the lowest number in the range was calculated, as it was more likely that the minimum rather than maximum number of participants were analysed. Finally, it was noted if data for children was provided individually (i.e., listed scores/data for each child) or together (aggregated scores for the intervention/comparator groups). If enough studies were similar and had individual-level data, it would potentially be possible to conduct an individual participant data (IPD) review. IPDs are when similar individuals from a number of studies are re-analysed together to offset bias from variable quality in reporting, and provide a more detailed analysis than when combining separate study effect sizes (Tierney et al., 2021). If individual data was not provided, then studies were considered to have aggregate group data only.

Participant information items

Participant information items focused on two pieces of participant information. First, the labels/diagnoses of the groups were noted. For example, children could either have a diagnostic label given by the author or via a prior diagnosis (e.g., DLD, SLI), or based on a more general label (e.g., ‘oral language delay’). This was to provide context for the types of children receiving intervention. Second, what authors reported regarding the child and social factors (of interest) for participants was extracted. This was to inform the implications of the findings when examined (i.e., what samples looks like, who these findings apply to, and does the ‘third variable’ include all relevant subgroups). It also helps understand what factors are considered important in samples, but which may or may not be considered for ‘third variable’ analysis. If data about participant information is unclear, it was assumed that this had not been a focus of the study but reported to indicate the type of sample.

Intervention and control information items

Data extracted included the trade name of the package if provided (e.g., *Talk Boost*) or a domain targeted by the intervention (e.g., vocabulary and reading), and the ‘intervention package’, which was completed for context and to help determine the homogeneity of the interventions and potential for meta-analysis. There are a number of elements which can fall under the umbrella term of ‘intervention package’. Law et al. (2004) and Marulis and Neuman (2013) outline this to consist of intervention type (e.g., intervention versus no intervention, alternative intervention, or treatment as usual), implementer (clinician, teaching staff, parent) and their demographic information, group size (i.e. if children were split into large or small groups, or 1:1 for the intervention), intervention fidelity (if the intervention is completed similarly for every child) and dosage (the amount of intervention a child receives).

Implementer demographic information, and fidelity data were not extracted, and dosage was extracted to give more context to intervention studies rather than be a focus in analysis. Implementer demographic information was not extracted because the focus was on the individual’s child and social factors. Additionally, demographic information and intervention fidelity are complex topics which have not been widely explored and would likely benefit from an analysis not possible to cover adequately here. Dosage was also not covered thoroughly as there is on-going, in-depth research examining its effects that also could not feasibly be conducted within the current study (e.g. Frizelle et al., 2021a, 2021b). As seen in the research by Frizelle and colleagues, dosage is reported in many forms and not all aspects of dosage are reported systematically. Unless dosage is reported consistently, it would be difficult to integrate and compare simply in the current analysis. Also, to have an in-depth analysis of these areas will be beyond the current study. ‘Intervention package’ information such as type, implementer, and group size were selected to provide a context for who participated in intervention activities and also determine whether this influenced findings. The control type (e.g., treatment as usual, alternative intervention, no intervention) with an overview of its trade name (if applicable) and procedure were also extracted for possible considerations of meta-analysis and to determine the variety of interventions present. Additionally, a brief overview for each intervention element (and training if available) is provided to give an overall impression of the intervention. The intervention which was the focus of assessment was designated as the intervention while other groups regardless of having an alternative intervention or not were considered controls.

Outcome descriptions and results information items

Measure names and type (e.g., standardised versus author-created) were extracted (outcome description information). Measurement type was considered because intervention effect sizes (and significance) may be affected by whether the measurement is explicitly aligned to the intervention (e.g., author-created) or more generalised (e.g., standardised measures) (Bakker et al., 2019). The implications for findings may be different too (i.e., improvements seen are specific or general). The specific oral language skill being targeted by the interventions was also extracted and then discussed with a trained speech and language clinician (CMK) to confirm JT’s understanding was correct. Group pre- and post- means for each outcome (both adjusted and unadjusted by weighting, sample changes, or the ‘third variable’ analysis) were obtained if available to calculate effect sizes (specified further in the effect measures section 2.2.8). Measurement times (pre-test, post-intervention, follow-up assessments after initial post-intervention) were also noted to provide an idea as to how long intervention effects were present. In addition, justification for choosing specific factors for 'third variable' analyses and related analyses (e.g., group similarities) were extracted if available. This was to further understand the utility of these analyses, and why specific factors are chosen for consideration.

Child and social factors, and 'third variable' analyses items

Eight factors of interest were chosen for extraction as possible 'third variables' which may be analysed in the intervention studies. They were 1) initial language ability; 2) non-verbal IQ; 3) co-occurring disorders; 4) age; 5) assigned at birth gender; 6) socio-economic status; 7) adverse childhood experiences; and 8) multilingual status. However, adverse childhood experiences and multilingual status were dropped due to the unavailability of data (see later in this section). Selection of these factors were based on literature concerning their potential impact on oral language development and intervention response. This is discussed with respect to each factor in turn in the following sections.

Data extracted included key descriptors of the factors (groupings, averages, thresholds, measures etc.), type of 'third variable' analyses conducted (i.e., subgroup, correlation, covariate, moderation and mediation) and the purpose of 'third variable' analysis in relation to oral language outcomes. In addition, all relevant statistical results (i.e., relevant to the research aims and factors of interest) reported were extracted. Details about how the variables were included in statistical models (alone, with others) was also collected. If data concerning 'third variable' analyses for factors were missing or unclear, the analysis was considered to be 1) either not a focus of the study (i.e., not using the factor reported in the sample in a 'third variable' analysis), or 2) potentially be a study at high risk for bias (i.e., missing data, only completing 'third variable' analysis for some outcomes). This is explored further in the risk of bias within studies section (2.3.6).

Initial language ability and language profile

When considering children with language difficulties, there appears to be two aspects of their language difficulties where intervention efficacy may be affected. The first is the severity of their language difficulty at the onset of the intervention. Studies suggest children with lower initial vocabulary and language comprehension scores demonstrate smaller growth in later oral language and literacy skills than higher achievers (Amorsen & Miller, 2017; Cabell et al., 2021; Green, 2021). As seen with Storkel et al. (2017), children with more severe vocabulary and phonological awareness gained less from the intervention compared to those with milder difficulties. Why children with more severe language difficulties respond less to intervention may be because their weaker language skills would make it harder for them to understand and engage in some steps/tasks related to language learning in the intervention (Storkel et al., 2017). This may in turn prevent them from gaining the maximum possible benefit from what is being taught.

Second, depending on reported language profile of difficulty children have (i.e., expressive versus receptive versus mixed difficulty), these may be more or less difficult to address in intervention. However, research is mixed. As noted in chapter 1, Boyle et al. (2007) found that older children (aged 6-11 years) with expressive difficulties gained more from interventions than those with mixed difficulties. In their later review, Boyle et al. (2010) suggested that children with language difficulties that have a receptive component (i.e., receptive alone, mixed) may have more difficulty developing language because they have more severe underlying difficulties in storing and processing language, and working memory compared to children with expressive language difficulties only. As such, it could be that children with receptive or mixed difficulties will gain less compared to children with expressive language profiles, as their cognitive disadvantages provide a barrier to their engagement in tasks promoting language learning. Counter to this hypothesis, some research has suggested that receptive language skills (like language comprehension) can be treated effectively when targeted using clinic-based interventions for older children (aged years 8+) and adolescents (Broomfield & Dodd, 2011; Ebbels et al., 2014, 2017). Therefore, if interventions are specifically targeting areas of difficulty, this more intensive focus on language learning may help fill in the gap seen in children with receptive or mixed difficulties. As such, it is unclear how children with different language profiles may respond to language interventions. Furthermore, much of the research was aimed at older children and adolescents. Therefore, it was important to examine in this review how language profiles may affect language intervention response, particularly for preschool-aged children.

Initial language information was extracted in terms of 1) its severity, and/or 2) the more general profile of the deficit. Initial language severity, language skill (e.g., expressive vocabulary, phonological awareness), and language profile data was extracted with relevance to 1) language measure (the measure itself and oral language skill), 2) the scores and severity (via reported SD, or by calculating the standard/percentile/scaled scores if provided); and 3) language profiles (reported expressive/receptive/mixed difficulties) indicated by the authors. The intervention and comparator groups in studies were required to have an average score of -1SD on at least one language measure, as this was the minimum considered threshold for evidence of impairment (Bishop, 2014). A score between -1SD and -1.5SD was labelled a mild language difficulty. Below -1.5SD was considered severe. Having two categories of scores - above or at and below -1.5SD – was chosen because children typically have a worse prognosis and outcomes at this level of impairment than children with milder difficulties (Verhoven & Segers, 2003). Additionally, a within-sample level of severity, where children were selected based on a specified cut-off of a measure's score (e.g., bottom 8 in a class, bottom 10% of the sample) could also be used to establish severity. This was included to better describe at risk of DLD samples.

Non-verbal IQ

Non-verbal IQ (NVIQ) is the ability to process, analyse and problem-solve information, which helps an individual understand their environment and act accordingly (Norbury et al., 2016). Assessments of NVIQ examine multiple cognitive domains, including visual processing, spatial perception, processing speed, attention and working memory (Deák, 2014; Grondhuis et al., 2018). Evidence regarding NVIQ is mixed. While Boyle et al. (2007) found that NVIQ does not impact language intervention response, more recent studies suggest that oral language development is related to NVIQ, and could impact language development and learning. While not every child with language difficulties performs poorly in NVIQ assessments (Bishop et al., 2017; Volkers, 2018); Griffiths et al. (2022) found evidence suggesting NVIQ and language development impact each other's rates of growth in children (aged between 7–13 years) who were typically developing or had language disorder. This is supported with evidence that has previously linked underlying cognitive processes to language outcomes for both typically developing children, and children with language difficulties. For example, Yim and Yang (2018) found that visual processing was weak in children with Specific Language Impairment (SLI) aged between 4 and 8. Willinger et al. (2019) found better processing speed predicted better language comprehension in older typically developing children (7 to 11 years), and Ebert (2021) found processing speed was slower in children with DLD (aged 6-8 years). Snijders et al. (2020) also observed better attention (measured at 18 months) promoted better language comprehension in toddlers (at 2 years old). Finally, Smolak et al. (2020) found evidence that visual–spatial sustained attention for DLD children (aged 7 years) was significantly lower when compared to typically developing peers. Findings also indicated relationships between sustained attention, working memory, and oral language for children with DLD, but not for the typically developing group. Furthermore, it has been reported that children with lower non-verbal cognitive scores have more challenges in engaging with learning activities (Alibali & Nathan, 2018). Therefore, children with language difficulties scoring lower on NVIQ assessments have more general cognitive difficulties that may not only provide barriers to engaging in learning tasks more generally, but also weaken children's language development which creates a barrier to understanding and engaging in tasks related to language learning. This may in turn prevent them from gaining the maximum possible benefit from what is being taught.

NVIQ was extracted with relevance to standardised measures as these are generally used for any study assessing cognitive ability. Therefore, SDs reported, or the standardised score was extracted. On a normal distribution, a score of 85 to 115 is deemed average, while scores at and below 84 (-1SD) and at and above 116 (+1SD) are considered below and above average respectively. Note that any study with children scoring on average below 70 on NVIQ measures would indicate a group with intellectual disabilities, and so were not included in this study (as this is considered a differentiating condition).

Co-occurring disorders

Co-occurring disorders describe the incidence of two or more disorders present for a child at the same time (McGrath et al., 2008). There is a heightened risk of children diagnosed with, or at risk of DLD also being at risk for or diagnosed with various other neurodevelopmental disorders (Bishop et al., 2017). Depending on how the internal mechanisms underlying co-occurring disorders relate to those underpinning language development/disorder, it could create a differing response to intervention compared to children with isolated language difficulties.

For example, Lewis et al. (2015) found that children aged between 4 and 6 years, who had speech sound disorder (SSD) and language impairment had poorer speech, language, and literacy outcomes than children with isolated SSD or isolated language impairment later in adolescence. Why children with poorer speech may also have poorer language outcomes is likely because weaker speech undermines the development of oral language by slowing phonological processing and ability to process and produce morphological structures (Haskill & Tyler, 2007; Lewis et al., 2015). This in turn creates a more severe language difficulty, which has been discussed above to create a barrier to children's engagement with language learning tasks (Storkel et al., 2017).

However, not all co-occurring disorders will likely function in this way. A set of prevalence studies by Redmond and colleagues' findings with a sample of older children with ADHD (2016; 2015) reported mixed evidence. In both of their studies, there were no apparent negative impacts of ADHD on oral language. But while one study (2016) found children with more severe ADHD showing better language outcomes, their other study indicated children with ADHD were more likely to be in speech and language services than children with isolated language difficulties. Why this occurs could be because their other difficulties (e.g., behaviour or attention) make children with ADHD 'stand out' to parents and educators, and so gain preferential access to services (Redmond et al., 2015). As such, ADHD may not have internal mechanisms that impact language development and intervention response, even if they are highlighted more to speech and language services.

With findings from these disorders in mind, it is important to consider if and how language intervention response may differ based on the type of co-occurring disorder. As the study was open to any disorder or difficulty which did not preclude a diagnosis of DLD, it was difficult to make specific predictions for every possible condition. But with the hypotheses made for SSD and ADHD in mind, it may be that some co-occurring difficulties may undermine certain language processes (e.g., speech and morphosyntax), and so are likely to have adverse differential impacts on intervention outcomes. In contrast, those which have unrelated underlying mechanisms to language will show equal gains from language intervention compared to those with related underlying mechanisms, and/or children with isolated learning disorders.

Outside of the findings reported, the primary focus of comorbidity research has been on prevalence (e.g. Eadie et al., 2015), and there appears to be little to no language intervention research for children with co-occurring difficulties. Still, it was possible that reporting and analyses around additional difficulties more generally (e.g., speech ability, behavioural difficulties) would be included in language intervention studies. Co-occurring disorders were extracted with relevance to any diagnosis allowable under the CATALISE criteria for DLD. The disorders specified by Bishop et al. (2017) as co-occurring disorders were “motor problems (developmental coordination disorder or DCD), reading and spelling problems (developmental dyslexia), speech problems, limitations of adaptive behaviour and/or behavioural, and emotional disorders” (p.1072). Associated assessments and scores for the co-occurring difficulty (e.g., articulation tests for speech difficulties) were extracted.

Age

As this study examined research which included children between the ages of 3 and 7 - a period of 4 years - age was included. Oral language development dramatically changes during early childhood (Honig, 2007; Jiang et al., 2018). But even within the early years, it is suggested that differential mutability (i.e., the ability to change) in language differs in relation to age. For example, evidence suggests that brain formation is mostly completed by the first 5 years of life, and so if language has been poor at the end of this ‘critical period’, then children will continue to perform poorly (Bylund, 2009; Pallier, 2007). As such, it could be hypothesised that the older that a child receives language intervention, the less benefit they receive compared to younger children.

However, recent research by McKean et al. (2015) and Taylor et al. (2013) instead suggests that children past 5 years (i.e. ages 7/8) can change and still benefit significantly from language intervention. Furthermore, they suggest that mutability in later preschool age can be related to a number of other individual and social factors.

Related to this, older children may benefit more in interventions because they are more experienced in educational tasks and generally more cognitively and socially developed, which allows them to access learning more easily (Cantalini-Williams et al., 2016). Therefore, it could be hypothesised that the developmental maturity older children have compared to younger children allows them to gain the maximum possible benefit from what is being taught. Ages were extracted with relevance to the sample’s average or range (if average was not available).

Gender assigned at birth

To date, studies examining children (and in the most part adults) typically report their populations of gender via a binary definition (male/female), and has not reached the point to explicitly describe samples outside of this definition (e.g., also including non-binary, genderfluid and agender identities). As such, gender for the current thesis will be considered in terms of a binary rather than a spectrum-based perspective. Males and females appear to demonstrate different levels of language difficulty and development. Language difficulties appear to be identified in boys more than girls (Talbot, 2020). Furthermore, studies examining typically developing children find that from as early as their first year of life, girls outperform boys in areas of expressive vocabulary, grammar, language comprehension and syntax language areas (Bouchard et al., 2009; Eriksson et al., 2012; Lange et al., 2016; Zambrana et al., 2012). If this is the case, it would suggest that if males have poorer language overall, then they may gain less from interventions (as outlined in the initial language ability and language profile section above). This is supported by the research from Boyle et al. (2007) found that girls gained more in receptive vocabulary compared to boys when receiving intervention from SLTs and SLT assistants. However, males are also shown to catch up once children enter school, where gaps in vocabulary, grammar and speech comprehension narrow by age 6, and performance in areas like vocabulary may be higher in males by the end of primary school (Lange et al., 2016; Rice & Hoffman, 2015), which may mean that differential intervention response may be less of an issue as children develop. However, it is unclear how this would translate to language interventions, and if gender differences in intervention response would be similar for different language outcomes. Gender assigned at birth was extracted with relevance to the number or proportion (if *n* numbers were not available) of male and female genders in the study samples.

Socio-economic status

Socioeconomic status (SES) is a complex multi-faceted construct which varies in definition and characterisation across studies (Bukodi & Goldthorpe, 2013; Letourneau et al., 2013). In general, SES measures refer to the material and socio-cultural aspects of a family, community or other social group, and can include specific attributes of the caregivers as a proxy for this (Erola et al., 2016; McLeod & McKinnon, 2007). Some of the most common facets of SES measured in studies are parental education level, marital status, employment status, household income, free school meals, household factors, and eligibility for subsidy or benefits (Ensminger & Fotherill, 2003; Lewis et al., 2016; Sarsour et al., 2011).

SES was chosen because evidence has demonstrated differences in language development depending on socio-economic group from as soon as children begin to use language. Fernald et al. (2013) found disparities in vocabulary and language processing efficiency at 18 months with a gap of 6 months in language ability when children were aged 2 years. In their reviews, Schwab and Lew-Williams (2016) and Pace et al. (2017) found research suggesting growth in language is also slower in children from lower compared to higher SES backgrounds. Additionally, high quality language input, opportunities and environments differed according to family resources (more available or present in higher SES families). Pace et al. (2017) also argue that poverty is more impactful for children under 5 than older children, which suggests that SES is particularly influential for preschool-aged children. These differences in exposure may then contribute to constant and even widening disparities in language development between children with different SES backgrounds (Neuman et al., 2018). Vocabulary development is not the only oral language skill which seems to be affected by SES, with grammar, narration and phonological development also showing large deficits of up to two years of language development in children from low SES backgrounds (Gardner-Neblett & Iruka, 2015; McDowell et al., 2007; Pace et al., 2017). As highlighted in chapter 1, poorer vocabulary intervention response has also been found for children from low SES backgrounds in the meta-analyses conducted by Marulis and Neuman (2010, 2013). Why this occurs could be that being socio-economically disadvantaged makes children less likely to be exposed to complex oral language, and language learning at home (Neuman et al., 2018). This in turn means children in lower socio-economic families have weaker language, which has been discussed above to create a barrier to children's engagement with language learning tasks (Storkel et al., 2017).

However, recent research by McKean et al. (2015, 2017) suggests that children aged between 4 and 11 from lower SES backgrounds and with low language may be able to catch up if language is monitored and targeted by intervention, and if their home learning and literacy environment is optimal. This is supported by Law et al. (2017), which suggested that children from lower SES demonstrated higher gains in language from parent-child book reading interventions. Therefore, it could be alternatively hypothesised that despite the gaps in language between children with low SES backgrounds and their more privileged peers, addressing resources (e.g., availability of more books and learning materials at home); and providing targeted interventions could bridge the large gap between different socio-economic groups. This in turn would create higher gains for children with lower socio-economic backgrounds.

As previously outlined, SES can be captured by a number of different factors. These different factors reflect different aspects of a family's resource, primarily falling in categories of material, educational, or social resources (Erola et al., 2016). It is best to include SES factors which fall under all three of these categories, because they appear to associate differently to language difficulties, and so only including one type and implying this represents all of SES could inaccurately describe the contribution of different SES variables and types (Erola et al., 2016; Vauhkonen et al., 2017). Therefore, an inclusive approach was taken such that, any factors or measures which fell under the broad umbrella of SES were extracted. Information extracted included SES aspects described (e.g., maternal education was examined and levels were based on qualifications obtained) and measured (e.g., name of government- or author-created measure) in the paper. Additionally, any SES subgroups (e.g., high, middle, low; no education, achieved high-school education, achieved university education) used in a study to describe participants/characterise the sample were noted. Measures could either be (inter)national or county/state-level government indexes (e.g. McLeod & McKinnon, 2007; Norbury et al., 2017), or research-developed measures which are typically smaller scale observations or questionnaires (sometimes based on larger scale measures; e.g. Sarsour et al., 2011).

Excluded from analysis: Adverse childhood experiences

Like SES, ACEs are complex and multi-faceted. They are defined as ‘potentially traumatic events that occur in childhood’ and also ‘aspects of the child’s environment that can undermine their sense of safety, stability, and bonding’ (Centre for Disease Control and Prevention, 2020). Examples of ACEs are (but not limited to) growing up in abuse, household violence, substance misuse, mental health issues, instability from parental separation or household member(s) being in prison (Centre for Disease Control and Prevention, 2020; Jimenez et al., 2016). It was unclear at the start of the project if there would have been many ACEs identified in papers explicitly or implicitly, and there was little to no literature discussing them in the field of speech and language. Only one study, Pears et al. (2014) examined a small number of children in foster care, and found that their phonological abilities were lower than expected. Therefore, it may be that children who experience ACEs are less likely to benefit from interventions if their language is on average lower (following initial language severity predictions outlined previously). In addition, there are other difficulties associated with ACEs such as more severe behavioural issues (e.g. Segal & Collin-Vézina, 2019). Behavioural issues in turn may be disruptive to their engagement with learning (Patalay et al., 2016), meaning they miss out on the maximum possible benefit from the intervention.

However, this factor was later excluded from analysis. This was because no studies described and analysed ACEs within their samples. Therefore, there was insufficient data to explore ACEs in the current study. Although there is an increasing awareness of ACEs, research seems to be conducted mostly in child health and abuse fields (e.g. Austin, 2018; Racine et al., 2018). Even within these fields, the focus is on mental and physical health and early development, which includes language, but not in the detail familiar to fields/studies which specifically focus on examining language development and intervention. To understand the full picture of language development and school readiness in the context of ACEs, researchers in the field of speech and language need to begin to understand and address adversities present in children’s lives. But, this can be challenging as ACEs are complicated and require careful consideration to research both methodologically and ethically (Assmusen et al., 2020).

Excluded from analysis: Multilingualism status

Multilingualism status refers to whether children speak and/or are exposed to one language (monolingual), or multiple languages (multilingual). Multilingual is used here as an umbrella term for children exposed to more than one language, as exposure can be very diverse in nature. For example, children can be learning multiple languages at once, or learn another after learning their mother tongue has started or is complete. The CATALASE statement for DLD notes that children have to present with language difficulties in all languages (Bishop et al., 2017). The initial plan was to extract data for ‘third variable’ analyses relating to multilingualism status, but this was decided against. This is because the literature currently does not characterise the diverse nature of multilingualism in interventions sufficiently. Specifically, understanding how multilingual status interacts with language development and disorder is challenging, and there are additional issues to consider for this that are not relevant to monolingual children (Crowe et al., 2021; Gathercole, 2018; Peña et al., 2020). For example, until recently many studies examining language difficulties excluded children if they were not monolingual, and instead multilingual children were treated more as a confound than a factor of interest (Marinis et al., 2017; Reilly et al., 2014a). In addition, assessments and interventions in studies typically focus on one language (Armon-Lotem et al., 2015; Engel de Abreu et al., 2013; Reilly, et al., 2014). As such, it would be difficult to extract data with the sufficient nuances needed to provide meaningful results for interventions and outcomes which only consider one language (as seen in the findings section, all studies included treated children in one language that was local to the country of study).

2.2.7. Study risk of bias assessment

The quality of studies was assessed using the Cochrane Risk of Bias Assessment Tool (Higgins et al., 2018). Six main areas of bias were assessed: 1) random sequence generation (how the study generates an allocation sequence for participants), 2) allocation concealment (if generated, could intervention allocations have been known before or during allocation), 3) performance bias (blinding participants and personnel from knowledge of the intervention the participant received), 4) detection bias (if outcome assessors were blinded to intervention allocation), 5) attrition bias and incomplete data (reporting attrition and exclusions, and reasons for this, plus any analyses conducted to determine if this affected results), and 6) selective reporting (whether all outcomes/data/analyses were reported). For each aspect of bias, a judgement of the possible risk of bias was made data extracted from the procedures, rated as ‘high risk’, ‘low risk’ and ‘unclear risk’ (if insufficient data is provided to make a judgement). For example, a study would be considered low risk in selective reporting if they reported all relevant analyses for each measure, while it would be high risk if they only reported the findings for some measures/ analyses. Another example for allocation concealment would be low risk if allocation to intervention was unknown by the research team and children, while high risk would be the opposite. For the sake of time and resources, half of the studies were judged by JT and the supervisory team to check that JT sufficiently judged risk of bias adequately, and then JT completed the rest of the judgements. Disagreements were resolved through discussion. Information outlining the judgements were tabulated by study and an overall description of the bias in all studies were presented narratively (appendix C).

2.2.8. Effect measures

The focus of the current review is to determine whether language intervention response differs by subgroups in child and social factors. Therefore, it was important to have effect sizes which reflected 1) initial intervention effects, and 2) ‘third variable’ analyses effects to compare differences. For initial intervention effects, the full group mean (i.e., full intervention group/ control group means) were used to calculate the effect size. For ‘third variable’ analyses effect sizes, subgroup means were used (e.g., calculating an effect size for both intervention v control for low SES children post-intervention means, and another for intervention v control for high SES children post-intervention means). The decision was made to calculate Hedges g effect sizes. This is because it is an educationally-based effect size rather than those suggested by Cohen (1992). Hedge's g effects are more suited to the types of interventions examined, and are different in nature to health-based interventions where Cohen is typically used. Specifically, samples within classrooms are highly heterogeneous, and how intervention study protocols and methodologies interact with such complexity will inevitably make effect sizes smaller than expected by Cohen's estimates (Bakker et al., 2019). Health interventions meanwhile are formulated with highly homogenous treatment procedures (i.e., surgical procedure, taking medication) and have more restrictive samples with less complex intervention components (e.g., do not usually teach a skill). As such, standardisation and randomisation are simpler to implement for these (Kraft, 2020). Therefore, educational/psychosocial interventions should not be compared to the same standards as health-based interventions (Kraft, 2020). Further, hedges g is more useful than Cohen's estimates in demonstrating effects in spite of a complex environment. This is because they make a correction for sample sizes and produces less upwards bias (i.e. for moderate and large effect sizes) than Cohen's d (Freeman et al., 1986; Hedges & Olkin, 1985). The magnitude of the effect sizes were interpreted with reference to the Education Endowment Foundation guidance (EEF; Coe et al., 2013), which reflect more educationally-based interventions. The EEF effect sizes are 0.01 to 0.18 for low, 0.26 to 0.44 for moderate, and 0.56 to 0.69 for high indicators of impact. Two types of effect sizes were taken: 1) The intervention-control group comparison effect sizes are when all children in each group are compared; and 2) the 'third variable' analyses also involve the influence of an additional factor in the intervention-control group comparison (e.g., both intervention and control group split by SES level and compared).

Data to calculate Hedges g effect sizes was extracted from study data provided (i.e., means and standard deviations of post-intervention data) and calculated for both initial and 'third variable' analysis effect sizes (where possible) using the following equation:

$$\frac{M_i - M_c}{\sqrt{\left(SD_c{}^2 + SD_i{}^2\right)/2}}$$

Where:

M is the mean

SD is the standard deviation

i is the number of participants in the intervention group

c is the number of participants in the control group

Effect sizes were calculated from studies' reported statistics and stored in an Excel document separately from the other extracted data, but were tabulated alongside the narrative data when synthesised.

As mentioned previously, the amount of evidence available was predicted to be minimal for examining how child and social factors affected intervention response. The function of the current systematic review is exploratory, attempting to draw in any data relating to child or social factors. As such, an inclusive approach was employed for extracting available findings/effect sizes. If multiple outcomes (i.e., expressive vocabulary, phonological awareness etc.) were measured, these were included. Further, if outcomes had multiple measures (i.e., multiple measures of expressive vocabulary), these were all included. Finally, studies which had multiple treatment arms (i.e., intervention group, control group 1, control group n, etc.) were included.

2.2.9. Synthesis methods

Synthesis was designed with reference to the PRISMA framework (synthesis decisions, data preparation, tabulation, and method). These are outlined below together with their rationale.

Synthesis decisions from available data

As a lot of qualitative data (e.g., participant and factor information) was used to help address the research objectives, a narrative review and synthesis was completed. In addition to a narrative review, the plan was to conduct a meta-analysis. However, this can only be appropriately conducted if individual studies are similar in nature. According to Borenstein et al. (2021), this includes having a homogenous design (QEs and RCTs would be grouped separately), similar interventions (similar components/ingredients), outcomes (had to measure the same skill) and study quality (no high and low biased studies together). For the current study, the same factor examined, and the same type of ‘third variable’ analysis was also considered. Otherwise, if studies were too heterogeneous, the summary result (i.e., overall effect size) would no longer be meaningful as findings would be subject to problems with accuracy, generalisability and bias (Akhter et al., 2019; Haidich 2014; Tugwell & Tovey, 2021). If studies looked sufficiently similar to the researcher, a further assessment of heterogeneity via statistical methods would have been employed (Lee, 2018). If data was similar, the aim would have been to also potentially pool individual participant data (IPD) rather than utilise aggregated group data if available.

When taking these requirements into account, meta-analysis and IPD were not possible due to substantial issues with heterogeneity in the identified studies. Study design, outcomes, and analyses varied widely between studies. As seen in the findings part of this chapter, studies were also variable in their quality, which made it inappropriate to group together. Furthermore, aside from heterogeneity issues, only one study provided individual level data, so IPD could not be completed. As such, a narrative synthesis with quantitative data as a support was conducted instead. Details of planned synthesis if a meta-analysis would have been possible are presented in Appendix B for interest.

As noted in the effect measures section (2.2.8), the plan was to calculate ‘third variable’ analyses effect sizes as well as initial intervention effect sizes for studies. Unfortunately, these could not be calculated for most studies. This was because the majority did not provide subgroup mean data or individual statistical results. Instead, a more general type of data (i.e., did the inclusion of a ‘third variable’ change the significance initial intervention result, and how) was extracted if effect sizes were not available.

Data preparation for synthesis

Effect sizes were calculated from the statistical values highlighted in the effect measures section (2.2.8). Data items (as described in the data items section, 2.2.6) were summarised for tabulation. If data required for effect sizes or narrative data was missing and could not be obtained, then effect sizes could not be calculated. It was not possible to impute this missing data into the final synthesis of findings due to the small amount of statistical data available. Microsoft Word was used to tabulate and synthesise the categorical and narrative data, which are presented in an APA format (American Psychological Association, 2021).

Data tabulation and display

Data was sorted and placed into tables, supported by narrative description. The narrative synthesis procedure is informed by Cochrane's narrative synthesis guidance (Ryan, 2013) and the PRISMA reporting items guidance (Page et al., 2021). The focus of the synthesis was the 'third variable' analyses results, while the results without 'third variable' analyses would only be used for comparison purposes.

2.2.10. Reporting bias assessment

For systematic reviews, it is important for publication bias of the collected studies to be considered as these could introduce bias into the overall synthesis (Song et al., 2012). As meta-analyses could not be performed, quantitative-based risk of bias tests could not be employed (e.g., funnel and forest plots) to examine publication bias. But as seen in the synthesis decisions section (2.2.9), the studies were very heterogeneous. In this case, publication bias is especially important to examine (Van Aert et al., 2019). Therefore, a narrative overview of publication bias was conducted (McGauran et al., 2010). Different publication biases (language and country of publications, date of publications, positive publication bias – reporting non-significant findings, and potential reference bias) was assessed. For missing data, authors were contacted and asked to supply it. If authors could not be contacted, or data could not be supplied, missing data and drop-outs would be noted for each individual study (Forero et al., 2019). To be transparent on where bias may occur, findings were explicitly labelled by which study they came from.

2.2.11. Certainty of evidence

As suggested by Schünemann et al. (2019), assessing the certainty of evidence is important for understanding how to interpret conclusions and develop recommendations based on the quality of evidence found. Specifically, they state assessing confidence in evidence is important to prevent extensive conclusions being drawn from findings that is based on little evidence. As such, the PRISMA guidelines suggest using GRADE (Grading of Recommendations Assessment, Development and Evaluation) to assess the certainty of evidence. The British Medical Journal (BMJ) publishing group (2017) outlines that GRADE assesses the certainty of evidence in five key ways, 1) risk of bias (overlaps with the Cochrane risk of bias already being completed), 2) imprecision (how effect estimates relate to the 95% confidence interval of the absolute effect), 3) inconsistency (the number of studies demonstrating consistent effect sizes), 4) indirectness (how studies directly compare interventions of interest to the participants of interest and report relevant and valid outcomes), and 5) publication bias (overlaps with what will be reported in the reporting bias assessment section). It has four ratings for certainty, ‘very low’, ‘low’, ‘moderate’, and ‘high’; with very low indicating the true effect is markedly different from the estimated effect, to high indicating that the true effect is similar to the estimated effect found. Due to the nature of GRADE, it is completed by considering the certainty of evidence for all studies together. Therefore, information outlining an overall judgement of synthesised studies were tabulated and an overall description is presented narratively.

2.3. Findings

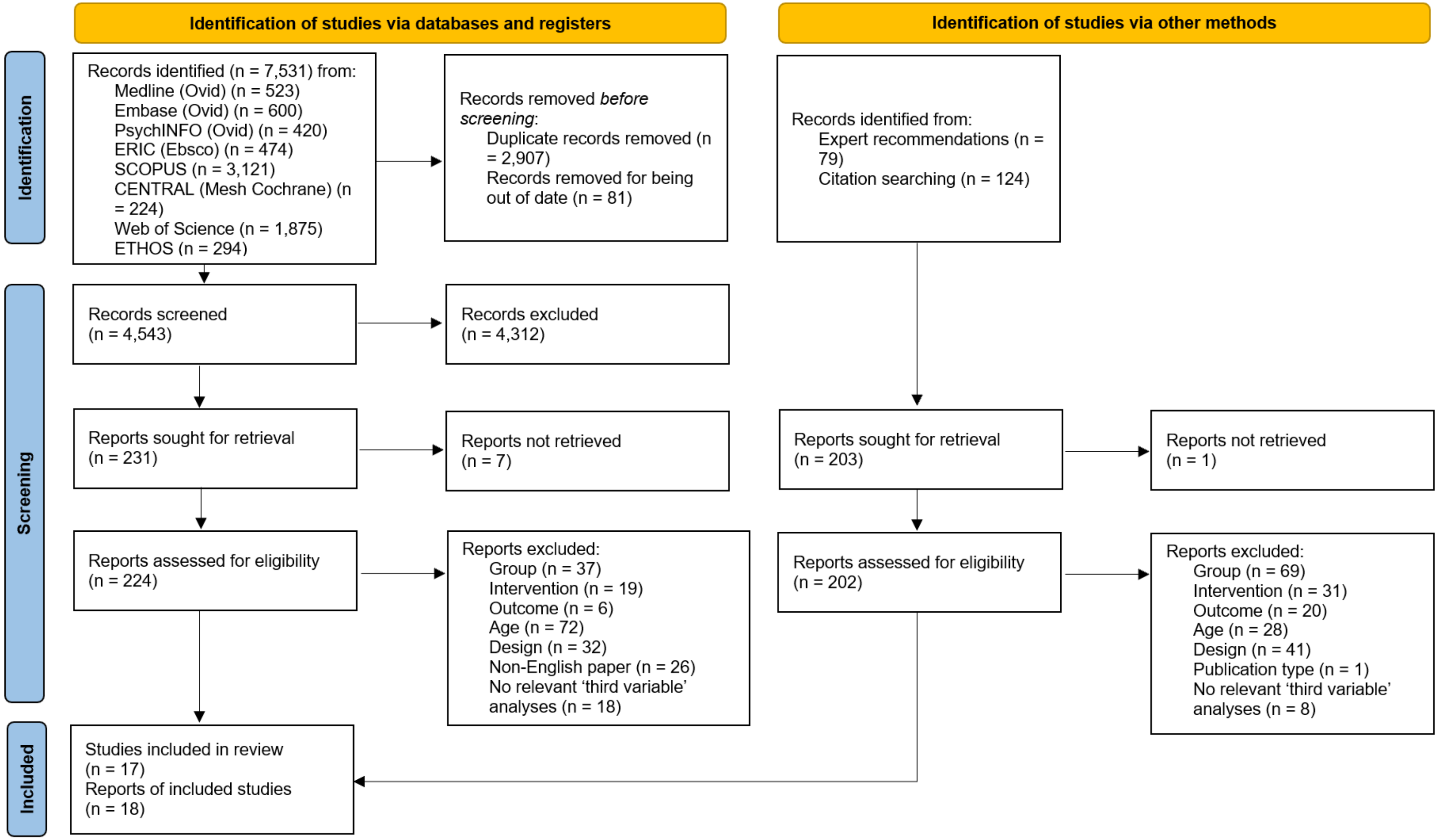
2.3.1. Study selection

The systematic review included any quasi-experimental studies or randomised controlled trials of language interventions, and included children aged 3-7 years who demonstrated language difficulties. Studies were also required to have completed ‘third variable’ analyses. 7,531 publications were found from selected databases and ETHOS. 2,907 papers were excluded for being duplicates, and 81 were excluded for being outside the date range. This left 4,543 papers and theses to be screened on title and abstract. At the end of the title and abstract screening, 4,312 papers were excluded from the database and ETHOS publications, leaving 231 remaining papers for full text screening. After full-text screening, 32 publications from these sources remained.

An additional 124 papers to the 32 above were identified as potential inclusions from hand-searching the bibliographies, and 79 were provided by expert recommendations (once duplications and out of date references were removed). Once full-text screening was completed, eight additional papers were included from bibliography searches and four from expert recommendation. It is important to note that typically experts suggested three to four papers, whilst one shared a reading list of 156 papers. Further, there was a high number of duplicates from the expert recommendations as was to be expected, explaining the unusual ratio of recommended to included papers seen here. However, identifying fewer studies at this stage demonstrated further validity of the database and ETHOS study acquisition and screening phases. No duplicates or out of date papers were noted for bibliography searched papers due to the nature of acquisition (i.e., JT only selected references which fell under criteria). At this stage, a total of 44 papers were included for consideration. Finally, JT checked the data relating to ‘third variables’ available from papers, and applied this final criterion. From this, 18 papers covering 17 studies were included for data extraction and analysis. When this process was completed, the final list was presented to CMK and JL, who considered that no additional papers were likely missing. A flow chart outlining the screening numbers created following PRISMA guidelines is provided in figure 2.7.

Each paper has been assigned a number (e.g., Aguilar et al. [1]) and will be referred to throughout using those numbers as a guide (see table 2.2). Wake et al.’s [15] 2013 and 2015 papers were analysed as one study because they were the same sample at 5 and 6 years old respectively. Age 5 outcomes were considered intermediate outcomes (tested after the year-long intervention) and age 6 outcomes were considered the definitive outcomes (tested around a year after the intervention). Smith-Lock et al.’s 2013 [12] and 2015 [13] papers were a feasibility study and a larger scale version of the same intervention, but were analysed separately due to differences in methodology, the participants and their child and social factors analysed.

Figure 2.7. PRISMA flow diagram for the systematic review



2.3.2. Study design information

There were ten QEs [1,3,4,6,7, 9,10,12,14, 16] and seven RCTs [2,5,8,11,13,15,17] included in the final synthesis. Nine studies were based in the US [1,4,6,7,9,10,14,16,17], four in the UK [2,3,5,11], three in Australia [12,13,15] and one in Germany [8]. All included samples completed English language interventions except for Motsch and Ulrich [8], where the intervention was in German. Papers were published between 2004-2018.

Across the included studies there were a total of 1,163 participants, with 581 children represented in the RCT studies, and 582 children represented in the QE studies. Sample sizes ranged between 18-180 participants (RCTs participant sample size = 31-180, mean 83; QEs sample size = 18-135, mean 58.2). The US studies generally had the smaller sample sizes (mean = 56.56), while the UK had the largest (mean = 107.5). Australian studies had a mix of small [12,13] and large [15] sample sizes (mean = 81.67), while the single German study [8] was the smallest (n=51). Although the top two largest samples were RCTs [e.g. 2,15], there were also relatively large sample sizes for QEs [e.g. 4, 10].

Table 2.2. Overview of studies by year, type, country and number of participants analysed

Study number, reference and year	Study type	Country of Study	N of participants analysed
[1] Aguilar, J. M., Plante, E., & Sandoval, M. (2018). Exemplar variability facilitates retention of word learning by children with specific language impairment. Language, Speech, and Hearing Services in Schools, 49(1), 72-84.	QE	USA	18
[2] Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., ... & Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. Journal of Child Psychology and Psychiatry, 49(4), 422-432.	RCT	UK	134-151
[3] Dockrell, J. E., Stuart, M., & King, D. (2010). Supporting early oral language skills for English language learners in inner city preschool provision. British Journal of Educational Psychology, 80(4), 497-515.	QE	UK	96
[4] Goldstein, H., Kelley, E., Greenwood, C., McCune, L., Carta, J., Atwater, J., ... & Spencer, T. (2016). Embedded instruction improves vocabulary learning during automated storybook reading among high-risk preschoolers. Journal of Speech, Language, and Hearing Research,59(3), 484-500.	QE	USA	105
[5] Haley, A., Hulme, C., Bowyer-Crane, C., Snowling, M. J., & Fricke, S. (2017). Oral language skills intervention in pre-school—a cautionary tale. International Journal of language and communication disorders, 52(1), 71-79.	RCT	UK	98

Study number, reference and year	Study type	Country of Study	N of participants analysed
[6] Justice, L. M., Kaderavek, J., Bowles, R., & Grimm, K. (2005). Language impairment, parent—child shared reading, and phonological awareness: a feasibility study. Topics in Early Childhood Special Education, 25(3), 143-156.	QE	USA	22
[7] Leonard, L. B., Camarata, S. M., Brown, B., & Camarata, M. N. (2004). Tense and agreement in the speech of children with specific language impairment. Journal of Speech, Language, and Hearing Research, 47, 1363-1379.	QE	USA	31
[8] Motsch, H. J., & Ulrich, T. (2012). Effects of the strategy therapy 'lexicon pirate'on lexical deficits in preschool age: A randomized controlled trial. Child Language Teaching and Therapy, 28(2), 159-175.	RCT	GER	51
[9] Phillips, B. M., Tabulda, G., Ingrole, S. A., Burris, P. W., Sedgwick, T. K., & Chen, S. (2016). Literate Language Intervention With High-Need Prekindergarten Children: A Randomized Trial. Journal of Speech, Language, and Hearing Research, 59(6), 1409-1420.	QE	USA	77
[10] Pollard-Durodola, S. D., Gonzalez, J. E., Simmons, D. C., Kwok, O., Taylor, A. B., Davis, M. J., ... & Simmons, L. (2011). The effects of an intensive shared book-reading intervention for preschool children at risk for vocabulary delay. Exceptional Children, 77(2), 161-183.	QE	USA	135
[11] Reeves, L., Hartshorne, M., Black, R., Atkinson, J., Baxter, A., & Pring, T. (2018). Early talk boost: A targeted intervention for three year old children with delayed language development. Child Language Teaching and Therapy, 34(1), 53-62.	RCT	UK	85

Study number, reference and year	Study type	Country of Study	N of participants analysed
[12] Smith-Lock, K. M., Leitao, S., Lambert, L., & Nickels, L. (2013). Effective intervention for expressive grammar in children with specific language impairment. International Journal of Language & Communication Disorders, 48(3), 265-282.	QE	AUS	34
[13] Smith-Lock, K. M., Leitão, S., Prior, P., & Nickels, L. (2015). The effectiveness of two grammar treatment procedures for children with SLI: A randomized clinical trial. Language, Speech, and Hearing Services in Schools, 46(4), 312-324.	RCT	AUS	31
[14] Van Kleeck, A., Vander Woude, J., & Hammett, L. (2006). Fostering literal and inferential language skills in Head Start preschoolers with language impairment using scripted book-sharing discussions. American Journal of Speech-Language Pathology, 15, 85-95.	QE	USA	30
[15] Wake, M., Tobin, S., Levickis, P., Gold, L., Ukoumunne, O. C., Zens, N., ... & Reilly, S. (2013). Randomized trial of a population-based, home-delivered intervention for preschool language delay. Pediatrics, 132(4), e895-e904. AND Wake, M., Levickis, P., Tobin, S., Gold, L., Ukoumunne, O. C., Goldfeld, S., ... & Reilly, S. (2015). Two-year outcomes of a population-based intervention for preschool language delay: an RCT. Pediatrics, 136(4), e838-e847.	RCTs	AUS	Age 5: 165-180 Age 6: 159-171
[16] Washington, K. N., Warr-Leeper, G., & Thomas-Stonell, N. (2011). Exploring the outcomes of a novel computer-assisted treatment program targeting expressive-grammar deficits in preschoolers with SLI. Journal of Communication Disorders, 44(3), 315-330.	QE	USA	34

Study number, reference and year	Study type	Country of Study	N of participants analysed
[17] Yoder, P. J., Molfese, D., & Gardner, E. (2011). Initial mean length of utterance predicts the relative efficacy of two grammatical treatments in preschoolers with specific language impairment. Journal of Speech, Language, and Hearing Research, 54, 1170–1181.	RCT	USA	57

Note. QE: Quasi-experimental study, RCT: Randomised control trial study.

2.3.3. Participants

The current section summarises the selected child and social participant factors extracted from papers. Almost all factors utilised in ‘third variable’ analyses were reported as part of participant information. Participant information for initial language ability, NVIQ, co-occurring disorders, age, gender assigned at birth, and socio-economic status will be detailed here, and flagged when relevant to the synthesis findings. A brief overview of participant information is shown in tables 2.3 and 2.4.

Initial language ability and language profile

To be included in the current review, samples had to be diagnosed or at risk of language difficulties. This subsection will highlight how participants were described in terms of diagnosis (e.g., SLI) and profiles (e.g., expressive/receptive/mixed), severity thresholds applied, measures used to describe language skills, and any use of additional but related criteria.

Diagnostic terms and profiles used

Nine studies [1, 6, 7, 8, 12, 13, 14, 16, 17] explicitly labelled children as SLI (specific language impairment) or LI (language impaired). This was based on a prior diagnosis, and/or decided through measurement scores. In eight studies [2, 3, 4, 5, 9, 10, 11, 15] participants did not have formal diagnoses, but were labelled as having poor ability, delay, or being at risk of diagnosable language difficulties. Only three samples described language profiles [6,8,15]. Children were described as having expressive language, and average receptive language [6], a vocabulary or word finding deficit [8], or expressive, receptive or mixed [15] difficulties. Motsch and Ulrich [8] also assigned diagnosis subgroups (SLI and Non-SLI) based on their achievement on their NVIQ measure (standard score of 85+ = SLI, 68-84 = non-SLI).

Language severity thresholds applied

Regardless of diagnosis, all studies utilised cut-point thresholds on one or more standardised measures of oral language. Participants in only two studies [1, 14] demonstrated more severe (below -1.5SD) language ability, while the rest were on average mild in comparison (-1SD to -1.5SD). As such, most interventions had samples of children with milder difficulties. Some studies included children with potentially more severe and close to average language abilities, as they expanded their thresholds (from average to around -2SD range) to include children who may not have fallen into their original inclusion criterion [2,4,11]. However, the average score of the sample was still below -1SD for Bowyer-Crane [2] and Goldstein [4], but unclear for Reeves [11]. However, Reeves' sample was labelled as 'at risk' of poor language, and so was included in synthesis.

Measures to describe language skills

All studies used at least one standardised measure to describe participants' language skills, but most used multiple measures [2, 3, 4, 7, 8, 9, 12, 13, 16, 17]. Participants were commonly identified to have general language difficulties [4,6,8,11,12,13,14] and/or difficulties with expressive morphosyntax (producing correct grammar-syntactic structures [1,2,7,9,13,16,17]). Participants' language skills were less commonly described in terms of: receptive vocabulary [3,4,5,9,10] and expressive vocabulary [2,3,5,15]; mixed morphosyntax and semantics (meaning as embedded in grammatical production, e.g., narratives) [3,9]; word knowledge (word definitions and description [4,15]); listening comprehension [5,9]; pragmatics [15]; and, phonological awareness [5]. This meant that the samples varied quite widely on what language difficulties participants had, although both expressive and receptive difficulties are represented in the selected studies.

Use of additional inclusion criteria alongside language ability

In three studies, children also had to show average scores or no impairment in specific language skills that were not the focus of intervention. These were for receptive language [16] and phonology [16, 17]. Inclusion by authors was not just based on measures of language skills, but in some cases included parental concern about language development [7], teacher selection [11], prior clinician-based diagnosis [12, 13] or author created measures [7, 16, 17]. Most studies selected participants recruited by the research team, but in four cases [2, 4, 5, 11], children were chosen as the *n* lowest scoring children in a classroom (e.g., in Bowyer-Crane et al. [2] it was the 8 lowest scoring children). In one case [4], there was designated cut-off and a proportion of children from each classroom were recruited; entry cut-off was relaxed to a milder language difficulty if not enough children met this criterion from a classroom.

Non-verbal IQ

NVIQ was an inclusion criterion for all nine studies which had Language Impairment (LI)/ Specific Language Impairment (SLI) samples [1, 6, 7, 8, 12, 13, 14, 16, 17]. Additional information about NVIQ was also provided in four samples with no specific diagnosis [2, 3, 5, 15]. Studies used either scaled [2,5], ability [3] or standard scores [1,6,7,8,13,14,16,17], and was unclear for two studies [12,15]. Almost all studies measuring NVIQ used only 1 assessment, except for Justice et al. [6] which used a different assessment for those below and above 4 years old, and Smith-Lock et al.'s studies [12,13] which accepted a range of NVIQ assessments completed by a prior diagnosing clinician. Motsch and Ulrich [8] also had two language profiles based on NVIQ score in their sample, labelling the participants as LI (below average NVIQ, standard score of 68-84) and SLI (average, standard score of 85+). Almost all studies required a standard/scaled score equivalent of 85/10 and above NVIQ. However, two of these studies [6, 17] allowed the inclusion of children scoring around -1SD below the mean (80-83 and above), and two others [1, 14] also included children provided they scored above the threshold for intellectual disability (i.e., standard score of 70). However, three of the four studies (except Justice et al. [6]) had samples which scored average NVIQ.

Co-occurring disorders

Co-occurring disorders present

Samples were not selected based on participants having any co-occurring disorders, but additional information on these was provided in eight studies [1, 2, 4, 6, 7, 10, 12, 15]. Additional information described children with speech-sound, speech intelligibility or articulation difficulties [1, 6, 12], behavioural scores [2, 15], children with 'independent educational plans' (unclear of what difficulties these were for, and if it related to their language difficulties) [4, 10] and children with actual or potential unspecified 'special' educational needs [7, 12]. These studies henceforth are categorised as comorbid speech difficulties (SSD and speech intelligibility/articulation [1, 6, 12]), behaviour [2, 15] and non-specific difficulties (children with independent educational plans and children with actual or potential unspecified 'special' educational needs [4, 7, 10, 12]).

Speech difficulties

Each study reporting comorbid speech difficulties had a sample with a moderate to high proportion of children with these. Only Aguilar et al. [1] provided a diagnosis of Speech Sound Disorder for their sample, while the others based speech difficulties on poor performance on the Goldman-Fristoe Test of Articulation (GFTA).

Behaviour

Behaviour scores in two studies were measured by the Strength and Difficulties Questionnaire (SDQ). Each study measuring behaviour had high proportions of children with behavioural difficulties, but their sample averages were under the SDQ threshold of high difficulties (14 and over). No specific diagnostic labels for children with high SDQ scores were assigned in either of the samples, but this may be because this is a screening measure.

Non-specific difficulties

Non-specific difficulties were not identified as particular disorders or given diagnostic labels, with studies only reporting proportions. For participants with independent educational plans, it was unclear if these were due to having other disorders, or were in place due to the language difficulty. In almost all cases, children with non-specific difficulties made up a very small proportion of the population, with the exception of Pollard-Durodola et al. [10], where independent education plans were present for 13% of the intervention group and 5.4% of the controls.

Exclusions

In the Smith-Lock et al. papers [12, 13], they explicitly did not include children if they had a diagnosis other than SLI, while Haley et al. [5] excluded children with identified special educational needs (but what this included was not specified). In three studies, samples were required to have age-appropriate articulation/speech skills [7, 16, 17].

Age

All studies reported ages within the sample, with five studies reporting range [1, 3, 8, 11, 15], and the rest reporting means. Participants in all samples were aged between 3;0 and 6;0. It is notable that few studies examined children at age six or seven. Other than Aguilar et al. [1] and Wake et al.’s [15] studies, no sample goes past five and a half years.

Gender assigned at birth

Gender assigned at birth was reported as a number or proportion in 14 of the 17 studies [4,11,17 did not]. Seven studies had similar levels of each gender (although there were always more males [2,3,5,8,9,10,14]), and seven had notably higher proportions of males [1,6,7,12,13,15,16].

Table 2.3. Overview of sample by child factors

Study	**Child factors as described in study samples**				
	Initial language abilities and label	**NVIQ**	**Co-occurring difficulties**	**Age**	**Gender assigned at birth**
[1] Aguilar et al.	Moderate language difficulty in expressive morphosyntax. Slightly below average in expressive and receptive vocabulary; labelled SLI	Around average of measure (standard score) Intervention mean: 94; Control mean: 100	Majority of sample have speech sound difficulties (14/18 children)	Range: 4;4-5;9 years	Mostly male sample (Males: 14; Females: 4, equal numbers per gender in intervention and control group)
[2] Bowyer-Crane et al.	Mild to moderate language difficulty in expressive morphosyntax and expressive vocabulary; labelled as language delayed and at risk of literacy problems	Below average of measure (scaled score) Children scored between 6 and 7	A large minority of sample have behavioural difficulties (intervention: 21.6%; control 22.4%). But the overall group averages within normal behaviour range (below 11 on SDQ difficulties total)	Mean: 4;09 years old	Equal males and females in sample (Males: 76; Females: 76, intervention: 40 males; control: 36 males)
[3] Dockrell et al.	Mild language difficulty in mixed morphosyntax and semantics, average ability to mild difficulties or	Below average of measure (standard	Not reported	Range: 3;0-5;0 years old	Roughly equal males and females in sample (Males: 65;

Study	Child factors as described in study samples				
	Initial language abilities and label	**NVIQ**	**Co-occurring difficulties**	**Age**	**Gender assigned at birth**
	expressive and receptive vocabulary; labelled as having low performance on oral language skills	score) Average score was 77.15			Females: 59, intervention: 40 males, control 1: 26 males; control 2: 15 males)
[4] Goldstein et al.	Mild language difficulty in general language, receptive vocabulary and word knowledge; labelled as at risk of reading or language disabilities	Not reported	Small proportion of sample have Individual Education Plans (intervention: 2.5%; control 5.1%; difficulties not specified)	Mean: 4;83 years old	Not reported
[5] Haley et al.	Mild language difficulty in expressive and receptive vocabulary, listening comprehension and phonological awareness; labelled as having poor oral language skills	Average (scaled score) mean score of 10	Children with special educational needs were excluded	Mean: 3;11 years old	Roughly equal males and females in sample (intervention males: 52%; control males: 55%)

Study	Child factors as described in study samples				
	Initial language abilities and label	**NVIQ**	**Co-occurring difficulties**	**Age**	**Gender assigned at birth**
[6] Justice et al.	Mild language difficulty in expressive language, average in receptive language; labelled SLI	Below average of measure (standard score) Children scored 80+, but unclear of mean	59% of sample (13/22 children) have speech difficulties (mean percentile rank GFTA scores of intervention: 29; control: 32)	Mean: 5;2 years old	Mostly male sample (Males: 18; Females: 4)
[7] Leonard et al.	Mild language difficulty in expressive morphosyntax, average to mild receptive vocabulary; labelled SLI	Scored average or above of measure (standard score of 85+)	Small proportion may have met criteria for other disabilities (numbers and disabilities not specified), excluded children with autism	Mean: 3;6 years old	Mostly male sample (Males: 25; Females: 6)
[8] Motsch & Ulrich	Mild to moderate general language difficulty; assigned SLI/LI depending on NVIQ	Both average (85+) and below average (68-84) NVIQ subgroups (standard score)	Not reported	Mean: 3;9 to 4;9 years old	Roughly equal males and females in sample (Males: 30; Females: 21, intervention males: 15; control males: 15)

Study	Child factors as described in study samples				
	Initial language abilities and label	**NVIQ**	**Co-occurring difficulties**	**Age**	**Gender assigned at birth**
[9] Phillips et al.	Mild language difficulty in expressive morphosyntax, mixed morphosyntax and semantics, receptive vocabulary and listening comprehension. About average expressive vocabulary; labelled as having low oral language skills	Not reported	Not reported	Mean: 4.53 years old	Roughly equal males and females in sample (Males: 45; Females: 37)
[10] Pollard-Durodola et al.	Mild receptive vocabulary language difficulty; labelled at risk for vocabulary delay	Not reported	Small proportion with Independent Education Plans (intervention: 13%; control: 5.4%; difficulties not specified)	Mean: 4;6 years old	Roughly equal males and females in sample (Males: 47%; Females: 53%)
[11] Reeves et al.	Mild/average general language difficulty; labelled as having delayed language development	Not reported	Not reported	Mean range: 3.48-3.53 years old	Not reported

Study	Child factors as described in study samples				
	Initial language abilities and label	**NVIQ**	**Co-occurring difficulties**	**Age**	**Gender assigned at birth**
[12] Smith-Lock et al.	Mild general language difficulty; labelled SLI	Within average NVIQ (various tests, no means provided)	Small proportion had special educational needs (intervention n: 1; control n: 3; difficulties not specified), minority with speech difficulties (6/34 children) (unclear if speech and special educational needs overlap), excluded if diagnoses not SLI	Mean: 5.1 years old	Mostly male sample (Males: 32; Females: 8, equal numbers per gender in intervention and control group)
[13] Smith-Lock et al.	Mild general language and expressive morphosyntax difficulty; labelled SLI	Scored average or above of measure (standard score of 85+)	Excluded if diagnoses not SLI	Mean: 5.1 years old	More males in sample (Males: 25; Females: 6)
[14] Van Kleeck et al.	Mild to moderate general language difficulty; labelled LI	Average of measure (standard score) Both intervention and	Not reported	Mean: 4;2 years old	Roughly equal males and females in

Study	Child factors as described in study samples				
	Initial language abilities and label	**NVIQ**	**Co-occurring difficulties**	**Age**	**Gender assigned at birth**
		control groups scored over 90			sample (Males: 17; Females: 13)
[15] Wake et al.	Mild language difficulty in expressive vocabulary, word knowledge and pragmatics; labelled as language delayed	Included if not demonstrating intellectual disability, but no measure or means for groups provided	Sample have typical behaviour on average (around 11 on SDQ difficulties total, means for intervention: 10.5; control: 9.4)	Mean Range: 4.1-4.2 at beginning, tested at 5 years (2013) and 6 years (2015) – not specified exact ages at testing	More males in sample (intervention females: 32%; control females: 36%)
[16] Washington et al.	Mild language difficulty in expressive morphology, average language for receptive vocabulary and general language; labelled SLI	Scored average or above of measure (standard score of 85+)	Average speech required (no score provided), and oro-motor or pervasive disorders excluded	Mean: 4;3 years old	More males in sample (Males: 27; Females: 7)
[17] Yoder et al.	Mild general language and expressive morphosyntax difficulty; labelled SLI	Scored average or above of measure (standard score)	Average speech required (standard score means of intervention: 90;	Mean: 3.6 years old	Not reported

Study	Child factors as described in study samples				
	Initial language abilities and label	NVIQ	Co-occurring difficulties	Age	Gender assigned at birth
		Intervention mean: 98; Control mean: 103	control: 91), excluded children with autism		

Socio-economic status

Thirteen of the seventeen studies [1, 2, 3, 4, 6, 9, 10, 11, 12, 13, 14, 15, 17] included information about SES as either an inclusion criterion [3, 9, 11, 12, 14], or as additional information [1, 2, 4, 6, 13, 15, 17], or both [10].

SES indicators present

The largest number of indicators of SES described was related to social capital. Participants were predominantly described by their geographical area (deprivation) data [1, 2, 3, 9, 10, 11, 12, 13, 14, 15], school funding/programmes attended (e.g. Head Start) [9,10,11], proportion of free school meal uptake [2,9,10], and presence of two parents in the household [6]. Educational levels (maternal and paternal or maternal only), by years in education [1] or level of qualification [6, 15, 17]) were also used to describe a notable portion of study samples. SES indicators of resource/income were used to describe participants the least, with only two studies reporting income [4] and parental occupational status [17].

Measures for SES indicators

SES for geographical area and proportion of free school meals was typically determined by government data and/or measures, while the other indicators were directly reported by the families. Occupational status was based on an economic measure in Yoder et al. [17], and it was unclear how income was reported in Goldstein et al. [4]. Children were typically selected due to their involvement with school funding/programmes due to their at-risk status.

Level of SES

Level of SES differed by study. Six described low SES samples [3, 4, 9, 10, 11, 14], four samples were labelled middle SES [1,6,15,17], two appeared to be mixed SES [2,13], and one was unclear [12]. Of the mixed SES samples, one study [2] had a larger minority of children from low SES backgrounds than expected via free school meals proportions, and/ or higher than expected numbers of low SES via area deprivation [i.e., 9,10]. Smith-Lock et al. [13] was designated as having mixed samples because they drew their samples from a variety of SES backgrounds. It should be noted that although Wake et al.'s [15] sample is labelled as average middle SES (due to mean geographical area and parental education data), families from lower and higher SES were also included.

Table 2.4. Overview of sample by socio-economic status

Study	Socio-economic status
[1] Aguilar et al.	Middle SES (maternal education mean - intervention: 14.3 years, control: 13.7 years; unclear SES for geographical area)
[2] Bowyer-Crane et al.	Likely mixed SES (higher than standard proportion of free school meals – intervention: 28.9%, control: 18.4%; unclear SES for geographical area).
[3] Dockrell et al.	Low SES (SES for geographical area = third most deprived borough in England)
[4] Goldstein et al.	Low SES (low income families, no further detail)
[5] Haley et al.	Not reported
[6] Justice et al.	Middle SES (via maternal and paternal education – 21 mothers and 18 fathers completed high-school, 14 mothers and 12 fathers completed university, 21/22 children had two-parent household)
[7] Leonard et al.	Not reported
[8] Motsch & Ulrich	Not reported
[9] Phillips et al.	Low SES (free or reduced school meals – 77-100% of children; SES for geographical area /School funding - schools with title I pre-k programs (typically used to support programmes in deprived areas/ families from low SES backgrounds, and children at risk of language difficulties)
[10] Pollard-Durodola et al.	Low SES (SES for geographical area /School programmes – selected schools with a high proportion of students from low SES backgrounds; free or reduced school meals – all children had this)

Study	Socio-economic status
[11] Reeves et al.	Low SES (SES for geographical area / School programmes – nurseries in socially deprived areas of North and NE England)
[12] Smith-Lock et al.	Unclear – but same SES (SES for geographical area)
[13] Smith-Lock et al.	Mixed SES (via geographical area – SEIFA (Socio-Economic Indexes for Areas) deprivation scores between 24th and 98th percentile)
[14] Van Kleeck et al.	Low SES (SES for geographical area /School programmes – All enrolled in Head Start preschool programmes)
[15] Wake et al.	Mixed/Middle SES (SES for geographical area - SEIFA deprivation scores average 1001 and 994 for intervention and control groups respectively; maternal education - did not complete school intervention: 30%, control: 26%; completed School intervention: 46%, control: 53%; obtained degree/ postgraduate qualification intervention: 24%, control: 21%)
[16] Washington et al.	Not reported
[17] Yoder et al.	Middle SES (parental education – average score indicates at least 4 years in university; parental occupational status – above average scores of 54 and 53 for intervention and control groups respectively (median is 29))

2.3.4. Intervention and control conditions

The current section summarises the intervention information extracted from papers. Intervention labels and types, targets, approach, context, agents of therapy and dosage will be briefly detailed here. In addition, a brief overview of the comparator control groups will be outlined. Further details for information provided in this section are highlighted in table 2.5.

Intervention labels and types

Ten studies had trade or specified names for their interventions [2,3,4,5,6,10,11,12,13,17]. None of the studies used the same interventions, except for the pair of studies by Smith-Lock et al. [12,13]. All interventions were generally facilitated by teaching/had a curriculum, and included activities and/or games. Over half of the studies also employed a storybook reading element [2,3,4,5,6,7,9,10,11,14].

Intervention targets

Interventions targeted selected vocabulary and word knowledge [1,3,4,5,8,10,11,15], individualised or general grammar targets [5,7,12,13,15,17], phonological awareness [2,6,15], narrative skills [5, 15], literal and inferential language skills [4, 14], sentence construction [11,16], listening skills [5,11], letter-sound knowledge [2], semantic and syntactic skills [9], preliteracy skills [15], and comprehension strategies [8].

Intervention approach

Of the 17 studies, five were explicit [6,7,12,13,17], 2 were implicit [2,11] and 10 were mixed [1,3,4,5,8,9,10,14,15,16] in their approach to teaching the assessed oral language skills.

Intervention context

The most common delivery context was via 1:1 therapy [1, 6, 7, 8, 14, 15, 16, 17]. Four studies delivered intervention in small groups of 2 to 4 children [3,4,5,9] and one study in larger groups of 5 or more [10]. Four studies had mixed group sizes, with one employing 1:1 and small group delivery [2], one utilising 1:1 and large groups [11], and two using both small and large groups [12, 13]. The majority of interventions were completed in a school setting/classroom [2,3,4,5,8,9,10,11,12,13], and less commonly completed at home [6,15]. For four studies [1,7,14,16], it was unclear where the intervention took place. However, is likely these were completed in a clinical setting because interventions were 1:1 and completed by clinicians or research associates. One study [17] confirmed intervention was completed in a university clinic.

Agents of therapy

The most common implementers were teaching staff [2, 3, 4, 5, 9, 10] and clinicians/speech and language therapists (SLTs) [1, 7, 8, 16, 17]. These agents of therapy were also utilised together in the Smith-Lock et al. studies [12, 13]. Parents [6] and research associates [14] were sole agents of therapy in one study each, and employed together in another study [15]. One final study utilised both teaching staff and parents [11]. Reporting the training of implementers varied, with six studies reporting no training details [1,4,7,8,16,17]. All but one [4] of those without training plans reported involved clinician/SLT implementers conducting 1:1 sessions. Otherwise, training sessions were reported for nine studies [2,3,4,6,9,10,11,14,15], lasting between an hour or half-day (e.g. Pollard-Durodola et al. [10, 11]) to several days (e.g. Dockrell et al. [2]). Six of these [2,5,6,9,14,15] also provided follow-up support, with five of these conducting observations to check fidelity [2,6,9,14,15]. Smith-Lock and colleagues [12,13] also reported providing a detailed manual about practice to implementers.

Intervention dosage

Dosage of interventions varied widely between studies. Intervention periods lasted between 3 to 26 weeks. Eight studies had intervention periods lasting at or below 10 weeks (3 weeks [1], 5 weeks [8], 8 weeks [12,13,14], 9 weeks [11], and 10 weeks [6,16]). Nine studies had intervention periods at or longer than 12 weeks (12 weeks [7,9,10], 15 weeks [3,5], 18 weeks [15], 20 weeks [2], 24 weeks [17], and 26 weeks [4]). The length of sessions generally fell between <10 to 15 minutes [1,3,4,14], 20 to 30 minutes [2,5,8,9,10,11,16,17] and 1 hour [12,13]. Two studies [6,15] had unclear session times, but this was likely due to being parent-implemented interventions taking place at home (and so were likely more flexible in timing). Leonard et al. [7] also did not report session times, but this is likely because the emphasis was on ensuring the children had a set number of exposures rather than keeping to a fixed session time. Sessions per week are detailed on table 2.5.

Comparison groups and intervention arms in analysis

For the comparison groups, the studies were split relatively equally into no treatment [3, 5, 8, 9, 10, 11, 12, 14, 15, 16] and alternative interventions [1, 2, 3, 4, 6, 7, 13, 16, 17] (note Dockrell et al. [3] and Washington et al. [16] had both types). In four studies [1, 3, 4, 13], alternative interventions were the same except for removing some intervention components. The remaining five studies [2, 6, 7, 16, 17] had an alternative or additional target, task or program of intervention (e.g., adding vocabulary building at the end of reading in Justice et al. [6], milieu language teaching as the alternative intervention in Yoder et al. [17]).

Table 2.5. Overview of intervention information, targeted outcomes, and 'third variable;' analyses for each study

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
[1] Aguilar et al.	**Name:** No official name **Procedure:** Taught unfamiliar target nouns (vocabulary) via presenting 3 varied object exemplars multiple times within themed activities (e.g building a child-sized robot and going on a pirate treasure hunt). **Dosage:** 3 sessions over 3 weeks (average session time 12 minutes).	Alternative intervention: Same but only presented with a single object exemplar.	1. Expressive vocabulary – (author created) "Generalised vocabulary measure"[2]	Initial language - expressive vocabulary (EVT-2); receptive vocabulary (PPVT-4) SES - maternal education Analysis: Correlation

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
	Implementer, location and mode: Clinician (1:1), unclear location, mixed. **Training:** Training information not specified.			
[2] Bowyer-Crane et al.	**Name:** "Jolly Phonics programme" **Procedure:** Taught letter-sound knowledge, phonological awareness (including articulatory awareness and sight word recognition) via blending and segmenting activities with integrated reading. **Dosage:** 20 weeks (20-30 minutes per session). **Implementer, location and mode:** Teaching staff (1:1 and small group), school, implicit. **Training:** 4 day training and fortnightly group tutorials by the research team and observed once teaching to assess treatment fidelity, when they also received feedback.	Alternative intervention: Received direct instruction to develop vocabulary, expressive language. grammatical competence and listening skills; encouraging independent speaking.	1. Expressive vocabulary – picture naming subtest of WPPSI-3[3] 2. Word knowledge – (author created) "specific vocabulary" [3] 3. Listening comprehension – NARA-2[3] 4. Mixed morphosyntax and semantics – Bus Story sentence length[3] 5. Mixed morphosyntax and semantics – Bus Story narrative skill[3] 6. Mixed morphosyntax and semantics ("expressive grammar")– APT[3]	Initial language – expressive vocabulary (picture naming, WPPSI-3), listening comprehension (NARA-2), mixed morphosyntax and semantics (Bus story sentence length; Bus story narrative skill; APT), phonological awareness (SIT) – all same initial language measure for respective outcome Age Gender assigned at birth

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
			7. Phonological awareness – SIT[3] 8. Phonological awareness – PAT[3] 9. Phonological awareness – TPA[3]	Co-occurring disorder - behaviour (SDQ total deviance) SES - Area deprivation score and child in receipt of free school meals Analysis: Covariate
[3] Dockrell et al.	**Name:** "Talking Time" **Procedure:** Taught targeted vocabulary via storybook reading and activities relating to story contents. **Dosage:** 2 sessions per week for 15 weeks (15 minutes per session) **Implementer, location and mode:** Teaching staff (small group), school, mixed.	C1 – Alternative intervention: Similar to I, but no training on how to talk with the children was provided. C2 – No intervention	1. Expressive Vocabulary - BAS Naming Vocabulary[3] 2. Receptive Vocabulary - BAS Verbal Comprehension[3] 3. Expressive Morphosyntax - GAP Sentence Repetition[3] 4. Mixed morphosyntax and semantics - Bus Story Information[3]	Initial language – expressive vocabulary (Naming vocabulary, BAS-2), receptive vocabulary (verbal comprehension, BAS-2), expressive morphosyntax (sentence repetition, GAP), mixed morphosyntax and semantics (Bus story

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
	Training: For intervention only, teachers were given specific information and training on certain ways of talking with children.		5. Mixed morphosyntax and semantics - Bus Story mean length of sentence[3]	information; Bus story mean length of sentence) – all same initial language measure for respective outcome NVIQ - BAS-2 Analysis: Covariate
[4] Goldstein et al.	**Name:** "The Story Friends Curriculum" **Procedure:** Took part in pre-recorded readings of storybooks, and were prompted to say words and definitions. Books were part of two series and 3 units that consisted of 9 instructional and 3 review books. 2 lessons were embedded on challenging vocabulary words and story questions, and 1 lesson on inferential story questions. **Dosage:** 3 sessions a week for 26 weeks (10-12 minutes per session). **Implementer, location and mode:** Teaching staff (small group), school, mixed.	Alternative intervention: The same intervention, but with no embedded lessons.	1. General language – CELF-P2[3] 2. Receptive vocabulary – PPVT-4[3] 3. Listening comprehension – (author created) – Assessment of Story Comprehension[3] 4. Word knowledge – (author created) – Unit Vocabulary Test[3]	Initial language – general language (CELF-P2), receptive vocabulary (PPVT-4), listening comprehension (author-created), word knowledge (author-created) - various combinations of the skills listed here

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
	Training: Training information not specified.			[+ classroom, treatment effects] Analysis: Covariate
[5] Haley et al.	**Name:** "The nursery Language4Reading (L4R) programme" **Procedure:** Taught vocabulary knowledge, narrative, grammar and listening skills via multisensory and narrative activities and interactive listening games in multiple contexts. **Dosage:** 3 sessions a week for 15 weeks (20 minutes per session). **Implementer, location and mode:** Teaching staff (small group), school, mixed. **Training:** Trained, provided with a detailed intervention manual and supported over intervention. Training day introducing the structure of language, its importance a child's academic experience and the programme details.	No treatment: Offered intervention according to need after school entry; however, this was not monitored by the research team and was implemented at the discretion of each participating school based on their interpretation of their children's post-test performance and the overall programme effectiveness.	1. Expressive vocabulary – CELF-P2 (expressive vocabulary)[3] 2. Receptive vocabulary – CELF-P2 (sentence structure)[3] 3. Phonological awareness – (author created) – "Alliteration matching"[3] 4. Word knowledge – (author created) – "word naming" [3] 5. Word knowledge – (author created) – "word definitions"[3] 6. Mixed morphosyntax and semantics – APT (information)[3]	Age Gender assigned at birth Analysis: Covariate

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
			7. Mixed morphosyntax and semantics – APT (grammar)[3] 8. Listening comprehension – Adapted YARC listening comprehension[3]	
[6] Justice et al.	**Name:** Phonological awareness-based storybook intervention **Procedure:** Completed multiple storybook readings (active involvement) a week with their parents which had both rhyming and narrative picture based books. Both a rhyme and alliteration task at the end of each storybook reading session was then completed. **Dosage:** 4 sessions a week for 10 weeks (unclear session length). **Implementer, location and mode:** Parent (1:1), home, direct. **Training:** Parents introduced to the book-reading intervention and tasks. Trained to engage in the	Alternative intervention: The same intervention but with vocabulary building tasks at the end.	1. Phonological awareness – (author created) - "Rhyme detection and production composite"[3] 2. Phonological awareness – (author created) - "Alliteration detection and production composite"[3]	Initial language – phonological awareness (rhyme detection and production, alliteration detection and production composites) – all same initial language measure for respective outcome, general language (TELD) – used for both outcomes Age

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
	two tasks at the end of reading. Provided reasons and modelling for tasks until they delivered with 100% accuracy. Trained to help children complete tasks via modelling the correct response, providing adequate wait time, and withdrawing support over time in response to children's progress.			Speech (GFTA) Analysis: Covariate
[7] Leonard et al.	**Name:** No official name **Procedure:** Taught third person singular - s targets via focused stimulation through storybook reading and acting the story out with toys, & with conversational recasting during play with toys and props. Clinicians engaged in recasting and were responsive to questions and requests by the child. **Dosage:** 4 sessions a week for 12 weeks (unclear session time). **Implementer, location and mode:** Clinician (1:1), unclear, direct. **Training:** Training information not specified.	Alternative intervention: The same procedure of the intervention condition was conducted, but children were taught auxiliary is/are/was.	1-4. Expressive morphosyntax targets (author created) - "3rd person singular –{s}" (intervention target), "is/are/was"[12]	Initial language - receptive vocabulary (PPVT-3) Analysis: Subgroup Age Analysis: Correlation
[8] Motsch & Ulrich	**Name:** No official name	No treatment	1. General language – AWST-R[23]	NVIQ (K-ABC-G non-verbal scale)

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
	Procedure: Taught semantic (e.g. to ask about word meanings) and lexical (e.g. to use phonological encoding) learning strategies for vocabulary via themed activities embedded with topics and phases. Also taught explicitly to ask questions relating to vocabulary they don't know. **Dosage:** 3 sessions per week for 5 weeks (30 minutes per session). **Implementer, location and mode:** Clinician (1:1), school, mixed. **Training:** Training information not specified.		2. Word knowledge - (author created) – "naming performance on trained words"	Analysis: Correlation NVIQ (K-ABC-G non-verbal scale) Age [phonological short term memory (K-ABC-G number recall)] Analysis: Covariate
[9] Phillips et al.	**Name:** No official name **Procedure:** Targeted semantic and syntactic skills (prepositional phrases, coordinating conjunctions, adverbial phrases, and negation) via structured language learning lesson unit plans. These consisted of an interactive adventure story, instruction on two story-embedded mental-state	No treatment	1. Expressive vocabulary – WJ-PV[34] 2. Receptive vocabulary – CELF-P2 sentence structure[34] 3. Word knowledge – (author created) [34]	Initial language - expressive vocabulary (WJ-PV), receptive vocabulary (sentence structure, CELF-P2), word knowledge (author created), and listening comprehension

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
	verb words, activities with manipulative props, and a picture game for review. **Dosage:** 4 sessions per week for 12 weeks (20 minutes per session). **Implementer, location and mode:** Teaching staff (small group), school, mixed. **Training:** Full-day training workshop and half-day booster session. Careful review of all lesson plans, materials, and intervention implementation procedures and opportunities to observe and discuss key features of sample lessons (via videos and live demonstrations) and practice with supervisor feedback. Provided ongoing professional development support throughout the intervention, which involved 1:1 consultation with the intervention designers and written implementation support guides specific to each of the units.		4. Listening comprehension - OWLS[34] 5. Listening comprehension - (author created)[34]	(OWLS; author-created measure) – all same initial language measure for respective outcome, plus other measures (CELF-P2 Concepts and Following Directions subtest, and CASL Syntax Construction subtest) Analysis: Covariate Initial language - expressive vocabulary (WJ-PV), receptive vocabulary (sentence structure, CELF-P2), word knowledge (author created), and listening comprehension (OWLS; author-created

Study	Intervention overview	Comparator(s)	Targeted outcomes	‘Third variables’ considered and analysis method
				measure) – all same initial language measure for respective outcome Analysis: Moderation
[10] Pollard-Durodola et al.	**Name:** “WORLD” **Procedure:** Taught vocabulary knowledge (words and meanings) via a curriculum utilising related sets of science-based vocabulary from informational and narrative text genres by lesson instruction, interactive book and informational text readings, and child-directed retelling tasks. **Dosage:** 5 sessions per week for 12 weeks (20 minutes per session). **Implementer, location and mode:** Teaching staff (large group), school, mixed. **Training:** Provided half-day training involving the rationale for intervention, materials, specific procedures, and the intervention architecture.	No treatment: Engaged in “practice-as-usual” shared book-reading activities and strategies determined by the teachers.	1. Expressive Vocabulary - EOWPVT[34] 2. Expressive Vocabulary - (author created) - RDEPVT[34] 3. Receptive Vocabulary – PPVT-3[34] 4. Receptive Vocabulary - (author created) - RDRPVT[34]	Initial language – expressive vocabulary (EOWPVT; author created measure), and receptive vocabulary (PPVT-3; author created measure) – all same initial language measure for respective outcome Gender assigned at birth Age Co-occurring disorder – non-specific difficulty

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
				(special educational status) [+ethnicity and multilingual status (Asian American)] Analyses: Covariate, Moderation
[11] Reeves et al.	**Name:** "Early Talk Boost" **Procedure:** Teaches attention and listening, learning words and building sentences via well evidenced language development practice and supported by a range of materials, a planning board, song cards, toys and a series of eight storybooks **Dosage:** 3 sessions a week for 9 weeks (20 minutes per session) **Implementer, location and mode:** Teaching staff (large group) and [Parent (1:1), school, implicit.	No treatment: Received the training after the reassessment of intervention children at post-intervention.	1 + 2. General language – PLS-4 expressive and receptive subtests 3. (1+2) General language – PLS-4 composite[1]	Initial language – general language (PLS-4 composite) –same initial language measure for respective outcome Analysis: Subgroup

Study	Intervention overview	Comparator(s)	Targeted outcomes	‘Third variables’ considered and analysis method
	Training: Attended a day training course raising awareness of children’s language development and improving ability to identify children with delayed language. Introduced the accompanying materials and to demonstrate the intervention sessions and the practitioner’s role in carrying them out. Parent training was a 1-hour workshop introducing key features of contingent behaviour when looking at book.			
[12] Smith-Lock et al.	**Name:** Expressive Grammar Programme **Procedure:** Taught expressive grammar via a focused grammar treatment programme for individually identified targets via general and specific grammar target teaching and activities. Groups were rotated to have 1 of each activity with TA, teacher and clinician. **Dosage:** 1 session per week for 8 weeks (60 minutes per session). **Implementer, location and mode:** Clinician and Teaching staff (Small-Large group), school, direct. **Training:** Manual of intervention provided.	No treatment: Focused on following directions and comprehension of prepositions.	1. Expressive morphosyntax - (author created) – “Grammar Elicitation Test”[1]	Co-occurring difficulty – Speech (GTFA) Analysis: Subgroup

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
[13] Smith-Lock et al.	**Name:** Expressive Grammar Programme **Procedure:** Cueing group: Taught expressive grammar via individually identified targets by general and specific grammar target teaching and activities and also included cueing (when a child made an error, the teacher/SLP followed a hierarchy of cues designed to elicit a correct answer). Groups were rotated to have 1 of each activity with TA, teacher and clinician. **Dosage:** 1 session per week for 8 weeks (60 minutes per session). **Implementer, location and mode:** Clinician and Teaching staff (Small & Large group), school, direct. **Training:** Manual provided (detailed activity plans, scripts and vocabulary).	Alternative intervention: Recasting group: Similar procedures, except for when following an error, the correct answer was provided to the child, but no attempt was made to have the child produce the target correctly.	1. Expressive morphosyntax - (author created) – "Grammar Elicitation Test"[3]	Initial language – expressive morphosyntax (author-created) – same initial language target for respective outcome (so different for the intervention and control groups) NVIQ (WNV-3) Analysis: Covariate
[14] Van Kleeck et al.	**Name:** No official name **Procedure:** Taught literal and inferential language skills via reading books and asked both literal (70%) and inferential (30%) questions about them	No treatment	1. General language – PLAI literal (levels I and II composite) [3]	Initial language – general language (PLAI literal; PLAI inferential composites), receptive

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
	via scripts embedded throughout the text. Also used scaffolding, prompts and appropriate response techniques. **Dosage:** 2 sessions per week for 8 weeks (15 minutes per session). **Implementer, location and mode:** Research Assistant (1:1), unclear, mixed. **Training:** Taught to ask the questions and provide the necessary prompts and responses in a standardised manner; and to extend and expand children's questions or comments related to the text. Videotaped sessions after every week were reviewed and discussed between implementer and author.		2. General language – PLAI inferential (levels III and IV composite) 3. Receptive vocabulary – PPVT-3	vocabulary (PPVT-3) – same initial language measure for respective outcome Analysis: Covariate
[15] Wake et al.	**Name:** No official name **Procedure:** Taught narrative skills, vocabulary, grammar, phonological awareness and preliteracy skills via sessions containing activities directed at the child; activities for parent and child together, with support from the language assistant; and activities for home practice	No treatment	1. Expressive vocabulary (age 5 and 6) - CELF-P2 Expressive Vocabulary[34] 2. Receptive vocabulary (age 6) – PPVT-4[34]	Initial language - expressive vocabulary (expressive vocabulary, CELF-P2), word knowledge (recalling sentences, CELF-P2) – used for word knowledge

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
	Dosage: 1 session per week for 16 weeks (unclear session time). **Implementer, location and mode:** Research Assistant and Parent (1:1), home, mixed. **Training:** 1 day group workshop, followed by individual 2-hour training with the supervising speech pathologist. Each assistant then observed supervising speech pathologist delivering at least 2 sessions and, once they had commenced delivering sessions independently, was observed by supervising speech pathologist on 2 occasions to ensure treatment fidelity. Two additional half-day group workshops on assessment and feedback were completed. The language assistants sought ongoing guidance from supervising speech pathologist as needed (~0.5 hours per week, per assistant). Activities for parent and child together, with support from the language assistant; and activities for home practice. Parents were asked to practice language-specific and storybook reading targets with their child during the week, and to keep diaries about each of these activities.		3. Word knowledge (age 5 and 6) - CELF-P2 recalling sentences[34] 4. Mixed morphosyntax and semantics (age 6) – Bus Story information[34] 5. Mixed morphosyntax and semantics (age 6) – Bus Story subordinate clauses[34] 6. Mixed morphosyntax and semantics (age 6)- Bus Story sentence length[34] 7. Phoneme awareness (age 5 and 6) - CTOPP[34] 8. Pragmatics (age 5 and 6) - CCC-2[34]	and expressive vocabulary outcomes at age 5, for all outcomes age 6, pragmatics (CCC-2) – only for pragmatics at age 6 Gender assigned at birth Age SES – maternal education Analysis: Covariate (findings represent a combination of these factors together) Age 5: Language profile - (expressive, receptive or mixed delay)

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
				Age 5 and 6: Initial language - expressive vocabulary (expressive vocabulary, CELF-P2), word knowledge (recalling sentences, CELF-P2) – used for word knowledge and expressive vocabulary outcomes at age 5, for all outcomes age 6 Age 5: NVIQ (unclear) Age 5 and 6: SES - maternal education Analysis: Moderation

Study	Intervention overview	Comparator(s)	Targeted outcomes	'Third variables' considered and analysis method
[16] Washington et al.	**Name:** No official name **Procedure:** Utilised a computer and software set-up to build up sentences, and provided opportunities for practice, reminders to use subjective pronouns, questioning and further probing for incorrect responses. **Dosage:** 1 session per week for 10 weeks (20 minutes per session). **Implementer, location and mode:** Clinician (1:1), unclear, direct. **Training:** Training information not specified.	C1 - Alternative intervention: Procedure completed with table top and tangible objects instead. C2: No treatment	1. Expressive Morphosyntax – SPELT-P[13] 2. Expressive Morphosyntax - DSS	Initial language – Expressive morphosyntax (SPELT-P) – same initial language measure for respective outcome Analysis: Subgroup (immediate post-intervention only) Initial language – Expressive morphosyntax (SPELT-P; DSS) – same initial language measure for respective outcome Analyses: Covariate (follow-up only for SPELT-P and immediate

Study	Intervention overview	Comparator(s)	Targeted outcomes	‘Third variables’ considered and analysis method
				post-intervention and follow-up for DSS)
[17] Yoder et al.	**Name:** Broad recast target intervention **Procedure:** Teaches grammatical targets via a system of child-directed play and graduated prompts for children to produce targeted language structures, with recasts, functional rewards, or verbal rewards to facilitate child language. **Dosage:** 3 sessions per week for 24 weeks (30 minutes per session). **Implementer, location and mode:** Clinician (1:1), clinic, direct. **Training:** Training information not specified.	Alternative intervention: Milieu language teaching: Similar prompts and methods to elicit three language targets for children based on their absence of utterances that are typically present in children with the target child’s MLU. Targets were replaced when children used three nominative examples of the structure in treatment sessions.	1. Expressive Morphosyntax– IPSyn[24]	NVIQ (LIPS-R) SES – maternal education Analyses: Correlation Initial language – expressive morphosyntax (mean length of utterances) Analysis: Moderation

Note. ‘Third variable’ analyses were conducted on outcomes as follows: [1] = subgroup; [2] = association; [3] = covariate; [4] = moderation. No number indicates no ‘third variable’ analyses were completed for that outcome. C1 = Control group 1; C2 = Control group 2. [] in ‘third variables’ considered and analysis method column = additional factors within the analysis, but not a focus for the current review.

Note. Standardised tests used for standard and ‘third variable ‘analyses per study: **[1]** EVT-2: Expressive Vocabulary Test–Second Edition (Williams, 2007); PPVT-4: Peabody Picture Vocabulary Test, 4th edition (Dunn & Dunn, 1997); **[2]** APT: The Action Picture Test (Renfrew, 2003); Bus Story [story sentence length, narrative skill] (Renfrew, 1991); NARA-2: Neale Analysis of Reading Ability, 2nd edition (Neale, 1997); PAT: Phonological Abilities Test (Muter, Hulme, & Snowling, 1997); SDQ: Strengths and Difficulties Questionnaire (Goodman, 1997) SIT: Sound Isolation Task (Hulme, Caravolas, Malkova, & Brigstocke, 2005); TPA: Test of Phonological Awareness (Hatcher, 2000); WPPSI-3: Wechsler Pre-School and Primary Scale of Intelligence 3rd edition (Wechsler, 2003); **[3]** BAS-2: British Ability Scales, 2nd edition (Elliott, Smith & McCulloch, 1997); GAPS: Grammar and Phonology Screening test (Gardner, Froud, McClelland & van der Lely, 2006); Bus story tasks (Renfrew Language Scales, Renfrew 1997); **[4]** CELF-P2: Clinical Evaluation of Language Fundamentals–

Preschool, 2nd edition (Wiig, Secord, & Semel, 2004); PPVT-4 = Peabody Picture Vocabulary Test, 4th Edition (Dunn & Dunn, 1997); **[5]** APT: The Action Picture Test (Renfrew, 2003); CELF-P2: Clinical Evaluation of Language Fundamentals - Preschool, 2nd edition (Wiig, Secord & Semel, 2006); YARC: York Assessment of Reading Comprehension (Snowling, Stothard, Clarke, Bowyer-Crane, Harrington, Truelove & Hulme, 2009); **[6]** GFTA-2: Goldman-Fristoe Test of Articulation first and second edition (Goldman & Fristoe, 1985, 2000); TELD: Test of Early Language Development (Hresko, Reid, & Hammill, 1991); **[7]** PPVT-3: Peabody Picture Vocabulary Test, 3rd edition (Dunn & Dunn, 1997); **[8]** AWST-R: Aktiver Wortschatz Test für 3- bis 5- jährige Kinder: Revised' (Kiese-Himmel, 2005); K-ABC-G: Kaufman Assessment Battery for Children, German edition (Melchers and Preuss, 1991); **[9]** CASL: Comprehensive Assessment of Spoken Language (Carrow-Woolfolk, 1999); CELF-P2: Clinical Evaluation of Language Fundamentals–Preschool, 2nd edition (Wiig, Secord, & Semel, 2004); OWLS: Oral and Written Language Scales (Carrow-Woolfolk, 1995); WJ-PV: Woodcock-Johnson III Tests of Cognitive Abilities - Picture vocabulary (Woodcock, McGrew, & Mather, 2001); **[10]** EOWPVT: Expressive One-Word Picture Vocabulary Test (Brownell, 2000); PPVT-3: Peabody Picture Vocabulary Test, third edition (Dunn & Dunn, 1997); **[11]** PLS-4: Pre-School Language Scale, 4th edition (Zimmerman et al., 2009); **[13]** WNV-3: Wechsler Nonverbal Scale of Ability, 3rd edition (Wechsler & Naglieri, 2006); **[14]** PLAI: Preschool Language Assessment Instrument (Blank, Rose & Berlin, 1987); PPVT-3: Peabody Picture Vocabulary Test, 3rd edition (Dunn & Dunn, 1997); **[15]** Bus story tasks (Renfrew Language Scales, Renfrew 1997); CCC-2: ; Children's Communication Checklist, 2nd edition (Bishop, 2003) CELF-P2: Clinical Evaluation of Language Fundamentals - Preschool, 2nd edition (Wiig, Secord & Semel, 2006); CTOPP: Comprehensive Test of Phonological Processing (Wagner, Torgesen & Rashotte, 1999); PPVT-4: Peabody Picture Vocabulary Test, 4th edition (Dunn & Dunn, 2007); **[16]** DSS: Developmental Sentence Scoring (Lee,1974); SPELT-P: Structured Photographic Expressive Language Test-Preschool (Werner & Kresheck, 1983); **[17]** IPSyn: Index of Productive Syntax (Scarborough, 1990).

Note. Author created measures used for standard and 'third variable 'analyses per study: **[1]** Generalised vocabulary measure; **[2]** Specific vocabulary; **[4]** Assessment of story comprehension, Unit vocabulary test; **[5]** Alliteration matching, Word naming; **[6]** Rhyme detection and production, Alliteration detection and production; **[8]** Naming performance of trained words; **[9]** Intervention-aligned assessment (IAA), Listening comprehension assessments; **[10]** Researcher developed expressive picture vocabulary test (RDEPVT), researcher developed receptive picture vocabulary test (RDRPVT); **[12,13]** Grammar Elicitation Test.

2.3.5. Outcomes

Oral language skills

The following outcomes were measured 1) general language (expressive and/or receptive language measured by omnibus tests [4,8,11,14]), 2) expressive vocabulary [1,2,3,5,9,10,15], 3) receptive vocabulary [3,4,5 9,10,14,15], 4) word knowledge [2,4,5,8, 9,15], 5) expressive morphosyntax [3,7,12,13,16,17], 6) mixed morphosyntax and semantics [2,3,5,15], 7) listening comprehension [2,4,5,9], 8) phonological awareness [2,5,6,15] and 9) pragmatics [15]. Details of the specific measures used are presented in table 2.5.

Measure types

Standardised measures were used for all outcome types, but authors used standardised measures 100% of the time for general language [4,8,11,14], mixed morphosyntax and semantics [2,3,5,15] and pragmatics [15]. In addition to standardised measures, author-created measures for expressive vocabulary [1], word knowledge [2,4,5,9], expressive morphosyntax [7,12,13], listening comprehension [4] and phonological awareness [5,6] were also used. Two studies used a mixture of both standardised and author-created measures for expressive and receptive vocabulary [10] and listening comprehension [9].

Post-intervention and follow-up data

Post-intervention times varied, with twelve studies only having one reported post intervention assessment period [3,4,5,6,7,9,10,11,12,14,17]. The period between the end of the intervention and the outcome assessments ranged from immediately after intervention to up to 2 [10] or 3 [5] weeks after. Six studies [1,2,8,13,15,16] also had a delayed follow-up outcome assessment, between 6 weeks and around a year after intervention. Only four of these six studies reported all measures at both times [1,8,13,16]. Only two studies [2, 15] completed relevant ‘third variable’ analyses for any follow-up findings (i.e., effect of covariates on intervention effects immediate and post-intervention in Bowyer-Crane et al. [2], age 5 and 6 ‘third variable’ analyses in Wake et al. [15]). Therefore, although other studies had full follow-up data, none of them conducted relevant ‘third variable’ analyses for their follow-up outcomes. As such, all but these two studies only present ‘third variable’ analyses of immediate post-intervention data.

2.3.6. Risk of bias within studies

The risk of bias of included papers were assessed using the Cochrane risk of bias assessment tool (Higgins et al., 2018). Each study was assigned if there was a high or low risk of bias, or an unclear risk of bias for methods of selection, performance, detection, attrition, and reporting. Critique in these aspects applies to the study as a whole. However, an additional aspect is explored in the current review due to the focus on ‘third variable’ analyses. This aspect is dubbed selective reporting of ‘third variable’ analyses, and considerations for bias are described and reported below. Judgements for all aspects are summarised in figure 2.8. The results reported here are for all eighteen papers covering seventeen different studies (Wake et al. [15] reporting in two papers). Evidence for judgements for each paper is presented in appendix C.

Random sequence generation and allocation concealment

In eight studies, it was unclear whether participants had been allocated randomly to groups [2, 3, 4, 6, 7, 8, 11, 13]. Researchers stated that participants had been randomly allocated in four studies [2, 6, 8, 11], but no further details were provided. Another study stated children had been randomised via a cluster randomised design, but did not state any details of the random sequence generation used [4]. For blinding of these allocations, only Bowyer-Crane et al. [2] specified that allocation was completed by a separate member of the team. The rest of these studies provided no information on this.

Random sequence generation and allocation blinding was not used in six studies [1, 9, 10, 12, 14, 16], and so they had a high risk of bias. In Aguilar et al. [1], children were assigned by their expressive vocabulary scores to ensure equal groups on this measure. In Phillips et al. [9], children who were randomly assigned to the treatment condition were then assigned to subgroups based on strategic decisions related to scheduling. Pollard-Durodola et al. [10] assigned children based on their enrolment to specific classrooms which had been randomly assigned to conditions, but this allocation was not described. Smith-Lock et al. [12] allocated children based on their school site and assigned a specific proportion of females to each group. In two studies, the children were assigned to a random group as soon as their permission form was returned [14, 16].

In three studies, randomisation was conducted using an algorithm or random number generator [5, 15, 17]; and so, presented a low risk of bias. In Wake et al. [15], allocation was concealed using sealed opaque envelopes. Yoder et al. [17] stated that the project director who enrolled participants was blind to treatment assignment at the time of enrolment. Haley et al. [5] stated the first author was initially blind to group membership, but this changed which implied the other authors were aware of groups initially and then the first author was later. Therefore, two [15,17] of the three studies were able to maintain a low risk of bias for allocation concealment throughout the study.

Performance bias blinding

All but one study [2] demonstrated either unclear or a high risk of performance bias. However, it is acknowledged that this type of bias is very difficult to mitigate in psychosocial-based interventions because their nature (e.g. implementers working directly with the children to improve language) makes it difficult to blind researchers, implementers and the participants (Banerjee et al., 2019; Grant et al., 2016).

Detection bias blinding

Outcome assessors were blinded in nine studies [1, 2, 3, 8, 9, 11, 13, 15, 16] and so demonstrated a low risk of bias. In seven of those studies [1, 2, 3, 8, 11, 15, 16], outcome assessors did not participate in intervention and/or were stated to be blind to allocation. In addition to staff not taking part in assignment or training, Phillips et al. [9] also anonymized pre- and post-intervention data and pooled this prior to double scoring by a blind scorer; Smith-Lock et al. [13] had different staff for pre- and post-intervention testing and scoring tests.

In seven studies [4, 5, 6, 7, 12, 14, 17], at least one of the research staff were aware of allocation and also administered some or all measures or transcribed and scored responses for children, meaning there was a high risk of bias. It should be noted that in Smith-Lock et al.'s [12] case, a comparison of the gain scores of the children tested by blind testers versus the children tested by a non-blind tester found no difference. In Pollard-Durodola et al. [10], no information relating to blinding outcomes was provided, and so bias was unclear.

Attrition bias incomplete data

A low risk of attrition bias was evident in five studies [1, 5, 8, 9, 12]. These studies had no missing outcome data or, where attrition had occurred, reasons were explicitly outlined and were judged to be unlikely to influence the true outcome. In Haley et al. [5], one child was excluded before randomisation occurred due to having a very severe language difficulty; they also provide a flow chart of attrition with reasons clearly stated and that appeared unlikely to affect true outcome (e.g., 5 lost due to moving schools). Motsch and Ulrich [8] also provided information on all dropouts, which occurred due to long-term illness or moving away. In Phillips et al. [9], the eight children missing at post-intervention in their study were divided equally between the treatment and control group in analysis. The five children in Smith-Lock et al. [12] who were dropped from the study had diagnoses which would preclude SLI (and so their inclusion may have impacted the validity of their results). The 'dropped' children appear to have taken part in the intervention (as it was within selected classrooms), but were not tested at any point, so removal from any analyses was not post-hoc.

Attrition information was not provided in six studies [6, 7, 10, 13, 14, 16] so the risk of bias was unclear. Unclear bias was also evident in three other studies [1,4,15]. Aguilar et al. [1] had one participant withdraw after the intervention phase, and replaced them with another who completed all study phases. However, it was unclear why this was. Goldstein et al. [4] found no accounted for attrition by recruiting more children, but if and why attrition occurred, or how it affected results was not explained further. Wake et al. [15] provided attrition information, but it was unclear from the information provided if this would influence the true outcome.

Four studies showed a high risk of attrition bias [2,3,11,17]. Although an attrition flow chart was provided in Bowyer-Crane et al. [2], certain events may have introduced bias (i.e., school withdrawing after allocation, 17 children being replaced following discussion with a teacher). This appeared due to their high performance on language measures, but there were no analyses provided to determine if this potentially affected results. Dockrell et al. [3] removed all of their monolingual participants due to having higher scores than those with multilingual (ELL) status, and it was evident that these groups differed significantly in language and NVIQ (this being the reasoning behind removing them). Reeves et al. [11] provided attrition information, but a notable number of nurseries (n=3) dropped out due to scheduling difficulties. There was no analysis as to how this would have impacted the true outcomes of the intervention. Finally, Yoder et al. [17] found that NVIQ and SES factors differed between drop-out and retained participants, meaning clinically relevant bias likely occurred in the outcome.

Selective reporting

Twelve of the seventeen studies were initially considered at low risk for selective reporting, reporting all pre-specified and expected outcomes. In the cases where there was a high risk, Bowyer-Crane et al. [2] did not provide all means and standard deviations (only providing a z-score bar chart for all outcomes that was difficult to decipher exact scores). They also did not provide all specified pre- and post- time points, and no exact numbers of participants for each outcome analyses were reported (just a range). Washington et al. [16] did not provide standard deviations in results. Dockrell et al. [3] did provide post-intervention data for the monolingual speakers together, but did not split the means by the intervention groups they were originally in, and as mentioned did not choose to include them in the final analysis. However, their reasoning for this was because the differences between both language groups was significant and there were uneven numbers of monolinguals in each group, which may have skewed and likely affected results greatly. Goldstein et al. [4] included school sites from two American states (Kansas and Ohio), but only reported initial group differences in language between children overall, and not by state (in contrast, comparisons for age and independent educational programs were completed for all children together, and by state). Justice et al. [6] did not provide mean and standard deviation scores for composites, and their z-score graph did not have exact numbers, so it was difficult to determine effect size.

Selective reporting of 'third variable' analyses

Developing an additional risk of bias criteria suitable for 'third variable' analyses

The Cochrane guidance for selective reporting focuses on ensuring that all stated outcomes undergo all reported analyses. But, as the current study has a specific focus on 'third variable' analysis, selective reporting is also considered here for 'third variable' analysis. It should be acknowledged however that this was used to enable an appropriate judgement on the level of confidence that can be had for studies' 'third variable' analyses. It does not reflect the study as a whole because studies were unlikely to have been set up with 'third variable' analyses explicitly in mind.

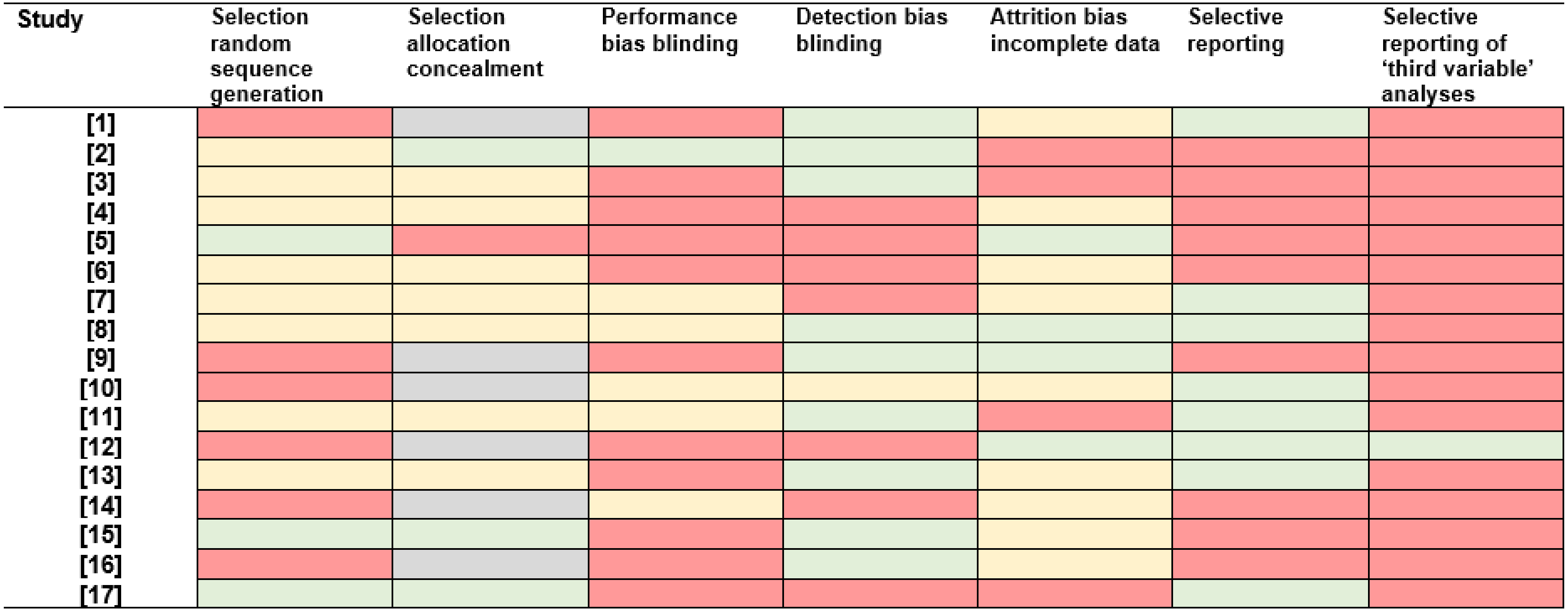
To determine selective reporting, three pieces of information were considered. These were considered because if it is unclear what the extent of the 'third variable' effect is, it may be over- or under-estimated. The first is the extent of missing statistical information for any 'third variable' analyses. High risk of bias could occur if information is 1) missing but interpretable (i.e., they state how 'third variable' effects outcome, but do not provide all statistics or adjusted and unadjusted means); or 2) missing and uninterpretable (no statement of how 'third variable' effects outcome or statistical information). While both are high risk as they do not provide full data, it is important to separate these as one can still provide some information, while the other does not. Second, it was determined whether 'third variable' analyses were completed for all outcomes (and if not, whether this was justified by the authors). The final aspect assessed was whether the same factors were used for each outcome where 'third variable' analyses were completed (e.g., if speech was analysed in subgroup analyses for each language outcome examined in the intervention study and if not, whether this was justified by the authors). These issues were considered to contribute to bias because their omission potentially distorts data by not finding out what their effects are.

Selective reporting results

All but one study [12] had a high risk of bias. There were also some uninterpretable findings from data reported in studies. This is defined as any ‘third variable’ analyses where authors do not report any narrative (e.g., stating a factor was a significant covariate) or statistical information (e.g., significance values) that could inform how child and social factors modified response to interventions. Therefore, some studies did not supply data that could be extracted, despite having completed ‘third variable’ analyses. There were two studies [5,14] where data could not be fully extracted. Both Haley et al. [5] and Van Kleeck et al. [14] reported conducting a covariate analysis, but explicit findings for these were not reported. Bowyer-crane et al. [2], Dockrell et al. [3], Phillips et al. [9], Smith-Lock et al. [13] and Washington et al. [16] also had covariate analyses (for some or all outcomes) which were uninterpretable for the same reasons.

Six studies [1,7,8,10,11,17] had at least some statistical information missing, but were interpretable because statistics provided could indicate an effect (e.g., a covariate significance value only) or studies stated how factors related to outcomes. Thirteen studies provided group comparison means that were adjusted [9,16] or unadjusted only [1,2,3,4,6,7,8,11,13,17] (or unclear [10]), so ‘third variable’ analyses effect sizes could not be compared to initial intervention effect sizes in most cases. In three studies [5,14,15], both unadjusted and adjusted data was provided, but not in a format that could be used to calculate effect sizes for comparisons. Eight studies [1,2,3,8,11,14,15,16] did not conduct ‘third variable’ analyses for all of their outcomes, and this was not justified. Five studies [2,4,8,14,15] did not assess the same factors in their analyses (e.g., typically adding or removing a factor as a covariate without clear justification).

Figure 2.8. Risk of bias overview by study



Note. Green indicates low risk, red indicates high risk and yellow indicates unclear risk of bias. Grey indicates not applicable, as selection random sequence generation was not completed.

2.3.7. Risk of bias across studies

As suggested by the PRISMA guidelines, an exploration was carried out into how the risk of bias presented in individual studies may affect conclusions made from the data. The majority of studies in the review exhibited either high or unclear level of bias for random sequencing generation and/or allocation to groups, with the exception of two studies [15, 17]. Performance bias for papers almost exclusively showed high or unclear bias with one exception [2]. Around 50% of papers [1, 2, 3, 8, 9, 11, 13, 15, 16] did however show a low level of detection bias. There was also a mix of low, high and unsure risk for attrition bias and selective reporting. It can be determined that overall, bias has been introduced because in many of the studies, participants, interventionists and outcome assessors would likely be aware of the different conditions in the study. Bias may have also been introduced to findings because it was unclear how participants not included in final analyses may have been different to those who were. Although general intervention study outcomes were reported in most studies, there was some missing data in a number of studies which was important to establish effect sizes, and analysing or reporting of ‘third variables’ for outcomes were either missing statistical data, omitted from an analysis, or uninterpretable. As such, caution should be advised in the interpretation of results due to bias being introduced in many of the included studies.

2.3.8. Result of synthesis

The current study explored how child and social factors affected language intervention response. This section outlines both of the objectives completed in order to address this, and is split into two parts. The first phase of analysis uses synthesised data to determine “*What participant factors are described in intervention studies for preschool language, and have been included in ‘third variable’ analyses, and why?”.* This provides an overview of which studies that described specific factors also utilised them in their ‘third variable’ analyses and why (if reported).

The second phase of analysis utilises the synthesised data to address *“What conclusions can be drawn regarding the impact of the identified child and social factors on preschool language intervention response?”.* It outlines results per each factor for each oral language outcome. These are presented per study, in addition to an overall synthesis of findings. Furthermore, an assessment of evidence confidence by analysis and study numbers is also presented alongside studies and overall.

Child and social factors both described for participants and included in 'third variable' analyses

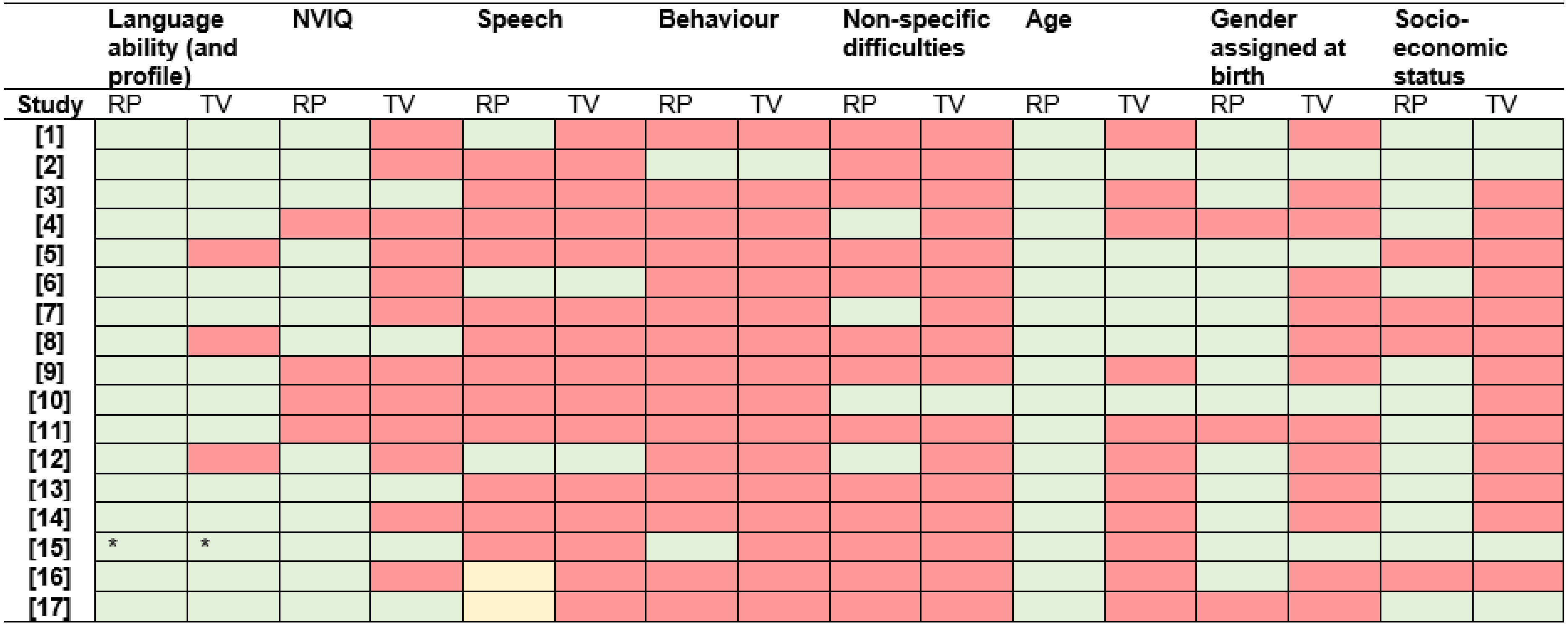
Choosing 'third variable' analyses

Child and social factors reported for participants were not always used in 'third variable' analyses. Some studies [4,5,13,14,16] explicitly reported decisions about analysing third variables based on pre-intervention group comparisons; specifically, whether the groups were judged to be similar/homogeneous. Many other studies grouped comparisons to ensure there was group equivalency for factors [1,4,5,6,8,10,13,14,16,17], indicating that authors acknowledged group differences in reported factors could potentially impact intervention results. However, choosing not to use reported participant factors at 'third variable' analysis stage was largely unexplained. Further, no study provided a detailed explanation for why factors needed to be equivalent in groups. Only two studies provided theoretical justifications for 'third variable' analyses [6,8], but these were both decided post-hoc.

Factors chosen

From all seventeen studies, only three [5,8,12] did not use initial language severity in 'third variable' analysis. Only Wake et al. [15] examined language profiles (expressive/receptive/mixed difficulties). Of the thirteen studies which described participants' NVIQ [1,2,3,5,6,7,8,12,13,14,15,16,17], only five [3,8,13,15,17] completed 'third variable' analyses for it. While all studies reported age, six studies [2,5,6,7,8,10] completed 'third variable' analyses' for it. Gender assigned at birth was used as a 'third variable' analyses for four [2,5,10,15] of the fourteen studies (all except [4,11,17]) that described it for participants. Two [6,12] of the three studies [1,6,12] which reported speech difficulties also utilised it in 'third variable' analyses. Behaviour was described in two studies [2,15], but only used in 'third variable' analyses for one [2]. Non-specific difficulties were described for participants in four studies [4,7,10,12], but it was only analysed as a 'third variable' in one [10]. Finally, four [1,2,15,17] of the thirteen studies [1,2,3,4,6,9,10,11,12,13,14,15,17] describing participants' SES included it for 'third variable' analyses.

Figure 2.9. Overview of factors described and used in 'third variable' analyses by study



Note. RP = Reported in participants, TV = used for 'third variable' analyses. Green indicates present, red indicates not present and yellow required average score for inclusion.

Note. * Indicates initial language and language profile were included in the study.

‘Third variable’ analyses and findings

‘Third variable’ analyses by child and social factor, analysis type and oral language outcome will be outlined here. A brief overview of findings are tabulated in tables 2.6 and 2.7, and tabulations by factors are provided in their respective sections. Findings highlighted red, yellow and green signify non-significant, mixed and significant findings respectively. Findings highlighted blue signify some aspect of the findings is unclear, and purple signifies both mixed and unclear findings. Uninterpretable data will not be synthesised alongside the other findings, but these have been detailed in the selective reporting of ‘third variables’ analyses section (2.3.6).

Splitting analyses by implications that can be made

When considering what could be concluded from the different ‘third variable’ analyses, it was decided that the findings should be split by 1) analyses which can determine how intervention response differs by different initial language/ language profile subgroups (subgroup and moderation analyses), and 2) analyses which only relate to the outcome growth/improvement (covariate and correlation analyses). While the former types can directly answer the research question; the latter types are unable to explain how different levels in the factor affect the outcome, but can relate directly to the gains made in the intervention. Therefore, even if the studies may not be able to indicate how child and social factors affect intervention response, they can be highlighted as possible areas to explore in future research (i.e., if significant, it would potentially be worth exploring as moderation analyses). For ease of reference, if there are more than four studies for a child or social factor, then tables will be split by 1) moderation and subgroup analyses; and 2) covariate and correlation analyses (e.g., tables 2.8 and 2.9 for initial language severity). Otherwise, if there are a small number of studies, findings will be placed together in one table (e.g., table 2.12 for speech). There were no studies which conducted mediation analyses with the chosen factors, and so this type of analysis has no findings presented.

Table 2.6. Overview of 'third variable' analyses exploring whether different levels in child and social factors cause differential intervention response by study

	Initial language		Language profile		NVIQ		Speech		Behaviour		Non-specific difficulties		Age		Gender assigned at birth		Socio-economic status	
	M	SG	M	SG	M	SG	M	SG	M	SG	M	SG	M	SG	M	SG	M	SG
GL	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
EV	3	-	1	-	1	-	-	-	-	-	1	-	1	-	1	-	1[1]	-
RV	3	-	-	-	-	-	-	-	-	-	1	-	1	-	1	-	1[1]	-
WK	2	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	1[1]	-
LC	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MMS	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1[1]	-
PRAG	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1[1]	-
PA	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1[1]	-
EM	1	2	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-

Note. Red indicates no effect, green indicates significant effect, yellow indicates mixed effect, and blue indicates unclear effect of characteristic on intervention response.

Note. Numbers in cells represents number of studies for result. M= moderation analysis, SG = subgroup analysis. GL = general language, EV = expressive vocabulary, RV = receptive vocabulary, WK = word knowledge, LC = listening comprehension; MMS = mixed morphosyntax and semantics, PRAG = pragmatics, PA = phonological awareness, EM = expressive morphosyntax. [1] = maternal/parental education, [2] = area deprivation and free school meals.

Table 2.7. Overview of 'third variable' analyses exploring how child and social factors relate to outcome growth by study

	Initial language		**NVIQ**		**Speech**		**Behaviour**		**Non-specific difficulties**		**Age**		**Gender assigned at birth**		**Socio-economic status**	
	Cov	Cor	Cov	Cor	Cov	Cor	Cov	Cor	Cov	Cor	Cov	Cor	Cov	Cor	Cov	Cor
GL	-	-	1	1	-	-	-	-	-	-	1	-	-	-	-	-
EV	4	1	1	-	-	-	1	-	1	-	1	-	1	-	1[1] 1[2]	1[1]
RV	3	-	1	-	-	-	-	-	1	-	1	-	1	-	1[1]	-
WK	2	-	-	-	-	-	1	-	-	-	-	-	1	-	1[1] 1[2]	-
LC	2	-	-	-	-	-	1	-	-	-	-	-	-	-	1[2]	-
MMS	2	-	-	-	-	-	1	-	-	-	-	-	1	-	1[1] 1[2]	-
PRAG	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-
PA	3	-	-	-	1	-	1	-	-	-	1	-	1	-	1[1] 1[2]	-
EM	2	-	1	2	-	-	-	-	-	-	-	1	-	-	-	1[1]

Note. Red indicates no effect, green indicates significant effect, yellow indicates mixed effect, blue indicates unclear effect, and purple indicates unclear and mixed effects of characteristic predicting/relating to outcome improvement/growth.

Note. Numbers in cells represents number of studies for result. Cov= covariate analysis, Cor = correlation analysis. GL = general language, EV = expressive vocabulary, RV = receptive vocabulary, WK = word knowledge, LC = listening comprehension; MMS = mixed morphosyntax and semantics, PRAG = pragmatics, PA = phonological awareness, EM = expressive morphosyntax. [1] = maternal/parental education, [2] = area deprivation and free school meals.

Confidence in evidence judgements

As part of the GRADE assessment domains (imprecision, indirectness, inconsistency), confidence in the evidence by analysis and number of studies was also examined by study. These were assigned the same levels as for the overall GRADE judgements (very low to high). These are shown in each outcome findings table (tables 2.8 - 2.19). An overall judgement per language outcome is also provided (in tables 2.8 – 2.19). Judgements were formulated according to 1) what type of 'third variable' analysis was conducted (i.e., highest confidence would be given to moderation analyses as they can demonstrate how different levels in factors create differential intervention outcomes; lowest confidence would be given to correlation as factors examined can only be said to associate with intervention growth), 2) the use of standardised or bespoke, author-created measures for outcomes (and measure for initial language), 3) If subgroups are specified, and whether they cover a good range of ability (i.e., they represent everyone that could potentially be examined), 4) whether findings are based on the factor alone or together with others, and 5) availability of statistical data to be able to calculate hedges g effect sizes, or data to report direction of effects. Author-created measures were ranked lower because their bespoke nature means that very specific aspects of a skill are assessed (e.g., vocabulary taught in the intervention), and so may not represent the full skill. It should be acknowledged that author-created measures may at times be the only alternative if standardised measures are not available – especially when languages other than English are involved. Nevertheless, these measures would be difficult to generalise and so are judged with less confidence. They are also not normed like standardised measures, so their validity and generalisability are lower. Where factors are combined, the study would not necessarily be ranked lower if individual data for each factor could be extracted. However, if effects could not be separated, this was ranked lower as it could not be determined what the individual effect of a specific factor was.

Initial language ability and language profile

Children's initial language ability severity was either split into subgroups, or measured along a continuum (depending on type of analysis used). In addition, there were three types of initial language ability used in ‘third variable’ analyses. These were 1) the same skill as the outcome [2,4,6,9,10,11,13,14,15,16] (e.g., how different severity of initial expressive vocabulary skills differentially impacts expressive vocabulary outcomes), 2) a different skill [1,3,4,6,9,7,15,17] (e.g., how different severity of receptive vocabulary differentially impacts expressive vocabulary outcomes), 3) the same skill but a different measure [1,17] (e.g., how initial expressive vocabulary measured by EVT-2 differentially impacts expressive vocabulary outcomes measured by vocabulary learned in the intervention). Although not necessarily ranked lower, different initial language skills to the outcome were noted. This was because this would have different implications to an analysis assessing the effect of the same measure/skill (e.g., initial receptive vocabulary and listening comprehension could have different underlying relationships with treated listening comprehension skills). This is defined separately from language profile, because the difficulties reported may not necessarily inform the full language profile of the child (i.e., children in a study could all have an initial vocabulary difficulty, but some of these children may have an expressive difficulty only, while others may have a more mixed difficulty). Only Wake et al. [15] considered language profile (expressive/ receptive/ mixed). Twelve studies had interpretable ‘third variable’ analyses for initial language [1,3,4,6,7,10,11,15,16,17].

Seven studies [7,9,10,11,15,16,17] had analyses which could determine how interventions are affected by different initial language/ language profile subgroups. These were split into subgroup analyses [7,11,16], and moderation analyses [9,10,15,17]. Subgroup analyses from one study [11] found that better initial general language meant children gained more benefit in the same skill from the intervention, but both higher and lower scoring groups (bottom 10% of scorers versus the rest of the sample) still benefitted at least moderately from the intervention. Subgroup analyses was completed in two studies [7,16] for expressive morphosyntax outcomes, but findings were mixed. However, the study where results were non-significant [7] examined the effect of initial receptive vocabulary (different language skill to outcome), and used an author-created measure for the outcome; while the study with significant findings [16] examined initial expressive morphosyntax (same language skill as outcome), and utilised a standardised measure at both points for the outcome. Leonard et al. [7] also only conducted subgroup analysis for the group’s target morpheme (–{s} for the intervention group, and auxiliaries *is/are/was* for the alternative intervention), so these findings should be considered in the context that both groups did not have ‘third variable’ analyses for the exact same outcome (although it was considered to reflect the same overall skill). In Washington et al. [16], better initial expressive morphosyntax predicted more benefit from the intervention, but all three groups (mean, +1SD of mean, -1SD of mean) still showed benefits (large effects for all subgroups).

Three studies completed moderation analyses for initial language skill. One study [9] examined listening comprehension, using two measures (one author created and one standardised). The findings were mixed, with results for the standardised measure being non-significant, and the author-created measure demonstrating that better initial word knowledge meant children benefitted less. Furthermore, only children scoring at the group mean and -1SD the group mean demonstrated a significant benefit from intervention. Mixed moderation findings were also found for the two studies [9,15] examining moderation for word knowledge. One study [9] found that better initial word knowledge meant more benefit (same skill measured), although all subgroups (mean, -1SD and +1SD) still benefitted to at least to a moderate extent. The other [15] found initial language (same and different language skills to the outcome) did not moderate intervention outcomes. The third moderation study [17] compared expressive morphosyntax outcomes for children who scored either below, or at and above 1.84 mean length utterances. However, findings were less clear. Although children in the lower scoring group did benefit from the intervention, children with higher initial morphosyntax scores could not be interpreted by the authors of the study. However, the study demonstrated that children with lower scores (below 1.84 mean length utterances) benefitted from the intervention. For expressive [9,10,15] and receptive [9,10,15] vocabulary, mixed morphosyntax and semantics [15], pragmatics [15] and phonological awareness [15], initial language was not a significant moderator. Expressive and receptive vocabulary were measured by a mixture of author-created and standardised measures in analyses, and two studies [9,10] assessed the same skills pre- and post intervention, while one [15] assessed different skills.

Confidence for evidence based on the number of studies was very low for all outcomes, with evidence being based on a maximum of between one and three studies depending on outcome and analysis type. Some of the outcomes were also inconsistent (e.g., listening comprehension, word knowledge and expressive morphosyntax), and this may be because the initial language measures used were different to the outcome measures, or significant findings (in moderation analyses) were based on bespoke author-created measures. Confidence in the analyses ranged from very low (mixed morphosyntax and semantics, pragmatics, phonological awareness) low (general language), to moderate (listening comprehension). Some analyses were mixed in their analyses confidence (expressive and receptive vocabulary = low to very low; word knowledge = very low to moderate; expressive morphosyntax = low to moderate).

Language profile (split into expressive/ receptive/ mixed difficulties) was examined in one study as a moderator [15], and did not affect intervention response for word knowledge and expressive vocabulary outcomes. Confidence in the analyses for the language profile analyses were rated moderate, but confidence in evidence based on number of studies was very low, due to being based on one study. Findings for initial language and language profile, and confidence judgements are presented in table 2.8.

Table 2.8. Overview of 'third variable' analyses exploring whether different levels in initial language severity/ language profile cause differential intervention response by study

Study	'Third variable' analysis method	Initial language measure	Outcome	Result	Confidence in analysis
[7] Leonard et al.	**Subgroup** – Split at and above, and below a score of 85	Receptive vocabulary (PPVT-3)	Expressive morphosyntax (author-created)	Non-significant difference on intervention response by subgroups	**Low:** Represents a comparison between lower scores and average +. Uses standardised measure for initial language, but has an author-created outcome. Would be better to split average and higher achievers and also compare this. Also requires an analysis of the same skill (can say initial ability in a different skill relates to outcome). Subgroup not as robust for establishing interactions as moderation. Effect sizes or direction of effect for each group could not be calculated with data reported.
[9] Phillips et al.	**Moderation** - Planned contrasts of intervention and control by subgroups of	Expressive vocabulary (WJ-PV)	Expressive vocabulary (WJ-PV)	Did not significantly moderate intervention response (only mean difference of groups effect size possible to calculate,	**Moderate:** Uses moderation analyses of the same skill. Has ability groups of high, mean and low scorers. Uses standardised measure. Effect sizes or direction of effect for each group could not be calculated with data reported.

Study	'Third variable' analysis method	Initial language measure	Outcome	Result	Confidence in analysis
	mean and ±1SD within the sample			non-significant group difference, g =.14)	
		Receptive vocabulary (sentence structure, CELF-P2)	Receptive vocabulary (sentence structure, CELF-P2)	Did not significantly moderate intervention response (only mean difference of groups effect size possible to calculate, non-significant group difference, g =.08)	**Moderate:** Uses moderation analyses of the same skill. Has ability groups of high, mean and low scorers. Uses standardised measure. Effect sizes or direction of effect for each group could not be calculated with data reported.
		Word knowledge (author created)	Word knowledge (author created)	Did significantly moderate intervention response (better initial word knowledge meant more benefit) Mean g =.88 -1SD g = .48 +1SD g = 1.27 (all significantly better for intervention group)	**Moderate:** Uses moderation analyses of the same skill. Has ability groups of high, mean and low scorers. Uses author-created measure. Effect sizes and direction calculated.
		Initial listening comprehension (same measure	Listening comprehension	Did not significantly moderate intervention response for OWLS measure	**Moderate (OWLS):** Uses moderation analyses of the same skill. Has ability groups of high, mean and low scorers. Uses

Study	'Third variable' analysis method	Initial language measure	Outcome	Result	Confidence in analysis
		for outcome, i.e., initial OWLS for OWLS outcome; initial author created measure for same outcome)	(OWLS, author-created measure)	(only mean difference of groups effect size possible to calculate, non-significant group difference, g =.31) Did significantly moderate intervention response (better initial word knowledge meant less benefit) Mean g =.79 -1SD g = 1.16 +1SD g = .35 (Only mean and -1SD significantly befitted in intervention compared to control group)	standardised measure. Effect sizes or direction of effect for each group could not be calculated with data reported. **Moderate (author created measure):** Uses moderation analyses of the same skill. Has ability groups of high, mean and low scorers. Uses author-created measure. Effect sizes or direction of effect for each group could not be calculated with data reported.
[10] Pollard-Durodola et al.	**Moderation** – using initial language as an interaction term	Expressive vocabulary (same measure for outcome, i.e., initial EOWPVT	Expressive vocabulary (EOWPVT, author created measure)	Did not significantly moderate intervention response.	**Low (EOWPVT):** Uses moderation analyses of the same skill. Uses standardised measure. No detail of any possible subgroups and effect sizes or direction of

Study	'Third variable' analysis method	Initial language measure	Outcome	Result	Confidence in analysis
	in multi-level model	for EOWPVT outcome; initial author created measure for same outcome)			effect for each group could not be calculated with data reported. **Very low (author created measure):** Uses moderation analyses of the same skill. Uses author created measure. No detail of any possible subgroups and Effect sizes or direction of effect for each group could not be calculated with data reported.
		Receptive vocabulary (same measure for outcome, i.e., initial PPVT-3 for PPVT-3 outcome; initial author created measure for same outcome)	Receptive vocabulary (PPVT-3, author created measure)	Did not significantly moderate intervention response	**Low (PPVT-3):** Uses moderation analyses of the same skill, and standardised measure. No detail of any possible subgroups and effect sizes or direction of effect for each group could not be calculated with data reported. **Very low (author created measure):** Uses moderation analyses of the same skill. Uses author created measure. No detail of any possible subgroups and effect sizes or direction of effect for each group could not be calculated with data reported.

Study	'Third variable' analysis method	Initial language measure	Outcome	Result	Confidence in analysis
[11] Reeves et al.	**Subgroup -** groups split between the bottom 10% scores and the rest	Initial general language (PLS-4)	General language (PLS-4)	Better language scoring group gained more than bottom 10% (study provided partial eta squares, hedges gs could not be calculated), although both groups significantly benefitted from intervention. Bottom 10% η2 = 0.107 Rest of sample = 0.132	**Low:** Represents a comparison between especially lower scores and the rest on the same skill. Would be better to have more subgroups (e.g., average and higher achievers separate) and give an indication of what the bottom 10% language scores actually are. Uses standardised measure. Subgroup not as robust for establishing interactions as moderation. Different effect sizes and direction of effect provided, but unclear how these would map to hedges g, as these could not be calculated with data reported.
[15] Wake et al.	**Moderation** – tests of interaction	Age 5: Language delay subgroup (expressive, receptive or mixed delay) Age 6: word knowledge	Expressive vocabulary (expressive vocabulary, CELF-P2)	Both language delay subgroup (expressive/ receptive/ mixed difficulties) and initial language did not significantly moderate intervention response.	**Moderate (language delay subgroup):** Uses moderation analyses. Well defined subgroups, based on standardised measures of CELF-P2 expressive vocabulary and/or recalling sentences scores more than 1.25 SD below the mean (so same and different language skill on outcomes measured). Effect sizes or

Study	‘Third variable’ analysis method	Initial language measure	Outcome	Result	Confidence in analysis
		(recalling sentences, CELF-P2) Age 6: expressive vocabulary (expressive vocabulary, CELF-P2)			direction of effect for each group could not be calculated with data reported. **Very low (recalling sentences, CELF-P2):** Uses moderation analyses for same and different skills (some outcomes are not examined with the same skill). Not specified cut-offs (only described as higher versus lower scores), and effect sizes or direction of effect for each group could not be calculated with data reported.
		Age 6: Initial language (higher versus lower scores)	Receptive vocabulary (PPVT-4)	Did not significantly moderate intervention response.	**Very low (expressive vocabulary, CELF-P2):** Uses moderation analyses. Used for same and different skills (some outcomes are not examined with the same skill). Not specified cut-offs (only described as higher versus lower scores), and effect sizes or direction of effect for each group could not be calculated with data reported.
		Age 5: Language delay subgroup (expressive, receptive or mixed delay) Age 6: Initial language (higher	Word knowledge (recalling sentences, CELF-P2)	Both language delay subgroup (expressive/ receptive/ mixed difficulties) and initial language did not significantly moderate intervention response.	

Study	'Third variable' analysis method	Initial language measure	Outcome	Result	Confidence in analysis
		versus lower scores)			
		Age 6: Initial language (higher versus lower scores)	Mixed morphosyntax and semantics (Bus Story information, Bus Story subordinate clauses, Bus Story sentence length)	Did not significantly moderate intervention response.	
		Age 6: Initial language (higher versus lower scores)	Pragmatics (CCC-2)	Did not significantly moderate intervention response.	
		Age 6: Initial language (higher versus lower scores)	Phonological awareness (CTOPP)	Did not significantly moderate intervention response.	
[16] Washington et al.	**Subgroup** – ANOVAs for intervention and	Expressive morphosyntax (SPELT-P)	Expressive morphosyntax (SPELT-P)	Mean and lower scoring subgroups gained more benefit than higher scorers	**Low:** Represents a comparison for high, mean and low scorers for same skill. Uses standardised measure. Subgroup not as

Study	'Third variable' analysis method	Initial language measure	Outcome	Result	Confidence in analysis
	control by subgroups of mean and ±1SD within the sample			(Study provided partial eta squares, hedges gs could not be calculated). Mean η2 =.69 -1SD η2 = .69 +1SD η2 = .25 (Only mean and -1SD significantly befitted in intervention compared to control group)	robust for establishing interactions as moderation. Different effect sizes and direction provided, but unclear how these would map to hedges g, as these could not be calculated with data reported.
[17] Yoder et al.	**Moderation** – Established a cut-point of ability (at and above 1.84 v below)	Expressive morphosyntax (author-created measure)	Expressive morphosyntax (IPSyn)	Children scoring below 1.84 significantly benefitted from intervention, but the result for children at and above group was uninterpretable.	**Very low:** Has statistically defined subgroups for moderation. Uses author-created measure for initial language. Effect sizes could not be calculated or direction of effect for each group unclear with data reported.
Overview of findings and confidence in evidence by outcome	**General language:** Significant difference in subgroups with different initial language on intervention outcome (1 study, low confidence in analysis). <u>Better initial general language meant more benefit from the intervention (moderate to large effects for both subgroups).</u> **Expressive vocabulary:** Non-significant moderating effects of initial language ability on intervention outcome (3 studies, between low and very low confidence in analyses). Non-significant moderating effects of language profile (1 study, moderate confidence in analysis).				

<table>
<tr><th>Study</th><th>‘Third variable’ analysis method</th><th>Initial language measure</th><th>Outcome</th><th>Result</th><th>Confidence in analysis</th></tr>
<tr><td></td><td colspan="5">Receptive vocabulary: Non-significant moderating effects of initial language ability on intervention outcome (3 studies, between low and very low confidence in analyses). Non-significant moderating effects of language profile (1 study, moderate confidence in analysis).

Word knowledge: Mixed moderating effects of initial language ability on intervention outcome (2 studies, significant study had moderate confidence in analysis, non-significant study had very low confidence in initial language analyses). Significant study = better initial word knowledge meant more benefit from the intervention (moderate to large effects for all subgroups). Non-significant moderating effects of language profile (1 study, moderate confidence in analysis).

Mixed morphosyntax and semantics: Non-significant moderating effects of initial language ability on intervention outcomes (1 study, 3 different measures, initial language skills were different to outcomes, very low confidence in analysis).

Listening comprehension: Mixed moderating effects of initial language ability on intervention outcome (1 study, standardised outcome measure non-significant, author-created measure significant, moderate confidence in analysis). Significant measure: better initial listening comprehension meant less benefit from the intervention (moderate to large effects subgroups).

Expressive morphosyntax: Mixed result for difference in subgroups with different initial language on intervention outcome (2 studies, non-significant study uses different initial language ability, while significant study uses the same measure/skill, low confidence in analysis for both significant and non-significant findings). Significant study = better initial expressive morphosyntax meant more benefit from the intervention (large effects for all subgroups).</td></tr>
</table>

Study	'Third variable' analysis method	Initial language measure	Outcome	Result	Confidence in analysis
	Partially unclear moderating effect of initial language ability on intervention outcome (1 study, one group uninterpretable findings, very low confidence in study analysis). Children benefitted from language intervention if they had poorer language scores, unsure of benefit if children had higher scores. **Pragmatics:** Non-significant moderating effect of initial language ability on intervention outcomes (1 study, initial language skills were different to outcomes, very low confidence in analysis). **Phonological awareness:** Non-significant moderating effect of initial language ability on intervention outcomes (1 study, initial language skills were different to outcomes, very low confidence in analysis). Confidence for analyses ranges between very low and moderate depending on outcome. Confidence based on study numbers for all outcomes is very low.				

Eight studies [1,2,3,4,6,10,13,15] had analyses which could relate initial language to intervention improvement/growth. These were split into correlation analyses [1], and covariate analyses [2, 3,4,6,10,13,15]. The correlation analyses [1] found initial language (standardised expressive and receptive vocabulary) did not relate to expressive vocabulary growth (author-created measure). For covariate analyses, initial language appeared to predict word knowledge outcomes in one study [4]. Goldstein et al. [4] also found a significant relationship between initial language and listening comprehension outcomes. Specifically, better initial language (mixture of same and different initial language skills) positively predicted outcomes (author created measures). However, it was unclear what the magnitude of these effects were. In Bowyer-Crane et al. [2] initial language (in combination of other factors – SES, age and gender assigned at birth) appeared to change their overall significant intervention effect to non-significant for one of their three measures of mixed morphosyntax and semantics, and a non-significant intervention effect to a significant one for listening comprehension (so in both cases it may not be initial language which influences the model, but SES). However, this was only seen at immediate post-intervention and not follow-up (which still demonstrated intervention effects) for mixed morphosyntax and semantics, and listening comprehension was only tested at immediate post-intervention so effects of initial language over time could not be determined. For expressive and receptive vocabulary, findings were also mixed. Dockrell et al. [3] found that initial expressive and receptive vocabulary significantly predicted outcomes in the same language skills, while Pollard-Durodola et al. [10] found that initial language did not significantly predict expressive or receptive vocabulary outcomes for either their author-created or standardised measures. Bowyer-Crane et al. [2] also found no changes to the significance of their initial group comparisons for expressive vocabulary outcomes. Expressive morphosyntax was also mixed, with one [3] of two studies [3,13] showing initial language to be a significant predictor. While all significant, it was unclear for all three outcomes in Dockrell et al. [3] what the magnitude or direction of these effects were. Finally, there were mixed findings for phonological awareness. Initial rhyme and alliteration score had significant and negative impacts on rhyme and alliteration growth respectively. Therefore, the higher children scored, the less benefit they received from the intervention. In the same study [6], initial general language had a significant and positive impact on rhyme growth, but not alliteration growth. Therefore, having better general language increases a child's gains in rhyme ability, while it has no effect on gains for their alliteration ability. Adjusting analyses via covariate analyses for multiple initial language skills in Wake et al. [15] also appeared to change the mean differences to a small extent positively for pragmatics, and negatively for expressive and receptive vocabulary, word knowledge, mixed morphosyntax and semantics, and phonological awareness, but it was unclear in any of these cases if this was significant.

Confidence for evidence in the correlation and covariate analyses based on the number of studies was very low for all outcomes, with evidence being based on a maximum of between one and four studies depending on outcome and analysis type. Some of the outcomes also had inconsistent findings (e.g., expressive and receptive vocabulary, listening comprehension, phonological awareness and expressive morphosyntax). Confidence in the all analyses were also very low. Findings and confidence judgements are presented in table 2.9.

Table 2.9. Overview of 'third variable' analyses exploring how initial language severity relates to outcome growth by study

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
[1] Aguilar et al.	Correlation	Expressive vocabulary (EVT-2) Receptive vocabulary (PPVT-4)	Expressive vocabulary (author created measure)	Non-significant relation to outcome growth	**Very low:** Cannot interact with intervention, can only relate to outcome growth for whole group (not different initial language ability subgroups). Used different (standardised) measure of same skill and measure of different skill to correlate with outcome. Outcome is author created. Effect size or direction of effect for each group could not be calculated with data reported.
[2] Bowyer-Crane et al.	Covariate Combined model: Two separate models with a) initial language of the same measure,	Expressive vocabulary (picture naming, WPPSI-3)	Expressive vocabulary (picture naming, WPPSI-3)	No effect on significance of initial group comparisons (measured at follow-up period only)	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Standardised measure of same skill. Effect of change in significance based on a combination of variables, rather than able to say the unique effect of

Study	‘Third variable’ analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
	age, gender assigned at birth and behaviour [SDQ total deviance]; b) the same but SES [Area deprivation score and child in receipt of free/reduced school meal uptake].				initial language. Effect size or direction of effect for each group could not be calculated with data reported.
		Listening comprehension (NARA-2)	Listening comprehension (NARA-2)	No effect on significance of initial group comparisons (measured at immediate post-intervention period only)	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Standardised measure of same skill. Effect of change in significance based on a combination of variables, rather than able to say the unique effect of initial language. Effect size or direction of effect for each group could not be calculated with data reported.
		Mixed morphosyntax and semantics	Mixed morphosyntax and semantics	Bus Story sentence length: No change to significance of initial group comparisons (measured at	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
		(same measure for outcome, e.g., initial Bus story sentence length for Bus story sentence length outcome)	(Bus Story sentence length, Bus Story narrative skill, APT)	immediate post-intervention and follow-up periods). Bus Story narrative skill: No change to significance of initial group comparisons (measured at immediate post-intervention period only). APT: No change when model a applied to initial comparisons, but model b changed significant difference between intervention and control groups (in favour of intervention group) to non-significant at immediate post-intervention; no change in significance at follow-up period.	initial language ability subgroups). Standardised measure of same skill. Effect of change in significance based on a combination of variables, rather than able to say the unique effect of initial language. Effect size or direction of effect for each group could not be calculated with data reported.
		Phonological awareness (SIT)	Phonological awareness (SIT)	No change when model a applied to initial comparisons, but model b changed non-significant difference	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
				between intervention and control groups to significant (in favour of intervention group, measured at immediate post-intervention period only).	initial language ability subgroups). Standardised measure of same skill. Effect of change in significance based on a combination of variables, rather than able to say the unique effect of initial language. Effect size or direction of effect for each group could not be calculated with data reported.
[3] Dockrell et al.	Covariate	Expressive vocabulary (naming vocabulary, BAS-2)	Expressive vocabulary (naming vocabulary, BAS-2)	Significant impact on outcome growth	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Standardised measure of same skill. Effect size or direction of effect for each group could not be calculated with data reported.
		Receptive vocabulary	Receptive vocabulary	Significant impact on outcome growth	**Very low:** Cannot interact with intervention, can only predict outcome

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
		(verbal comprehension, BAS-2)	(verbal comprehension, BAS-2)		growth for whole group (not different initial language ability subgroups). Standardised measure of same skill. Effect size or direction of effect for each group could not be calculated with data reported.
		Expressive morphosyntax (sentence repetition, GAP)	Expressive morphosyntax (sentence repetition, GAP)	Significant impact on outcome growth	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Standardised measure of same skill. Effect size or direction of effect for each group could not be calculated with data reported.
[4] Goldstein et al.	Covariate Combined model: Initial language –	Word knowledge (author created measure)	Word knowledge (author created measure)	Significant positive impact on outcome growth	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Multiple initial language measures

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
	general language (CELF-P2), receptive vocabulary (PPVT-4), listening comprehension (author-created), word knowledge (author-created) - various combinations of the skills listed here [+ classroom, treatment effects]	Listening comprehension (author created measure) General language (CELF-P2) Receptive vocabulary (PPVT-4)			entered into model. Mix of author-created and standardised predictors, author created outcome. Effect size or direction of effect for each group could not be calculated with data reported
		Listening comprehension (author created measure) Word knowledge (author created measure)	Listening comprehension (author created measure)	Significant positive impact on outcome growth	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Multiple initial language measures entered into model. Mix of author-created and standardised predictors, author created outcome. Effect size or direction of effect for each group

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
		General language (CELF-P2) Receptive vocabulary (PPVT-4)			could not be calculated with data reported
[6] Justice et al.	Covariate	Phonological awareness (same author created measure for outcome, i.e., initial rhyme for rhyme outcome; initial alliteration for alliteration outcome) General language (TELD)	Phonological awareness (author created measures of rhyme and alliteration)	Initial rhyme score had significant and negative impact on rhyme growth (higher initial score, less benefit). Initial alliteration score had significant and negative impact on alliteration growth (higher initial score, less benefit). General language had significant and positive impact on rhyme growth (higher initial score, more benefit), but not for alliteration growth	**Very low (rhyme and alliteration author-created measures):** Cannot interact with intervention, can only individually predict outcome growth for whole group (not different initial language ability subgroups). Multiple initial language measures and other factors entered into model. Mix of author-created and standardised predictors, with author created outcomes. Effect sizes for each group could not be calculated with data reported

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
[10] Pollard-Durodola et al.	Covariate	Expressive vocabulary (same measure for outcome, i.e., initial EOWPVT for EOWPVT outcome; initial author created measure for same outcome)	Expressive vocabulary (EOWPVT, author created measure)	Non-significant relation to outcome growth	**Very low (EOWPT):** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Standardised measure of same skill. Initial language measure and other factors entered into model together. Effect size or direction of effect for each group could not be calculated with data reported. **Very low (author-created measure):** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Author-created measure of same skill. Initial language measure and other factors entered into model together. Effect size or direction of effect for each

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
					group could not be calculated with data reported.
		Receptive vocabulary (same measure for outcome, i.e., initial PPVT-3 for PPVT-3 outcome; initial author created measure for same outcome)	Receptive vocabulary (PPVT-3, author created measure)	Non-significant relation to outcome growth	**Very low (PPVT-3):** Cannot interact with intervention, can only individually predict outcome growth for whole group (not different initial language ability subgroups). Standardised measure of same skill. Initial language measure and other factors entered into model together. Effect size or direction of effect for each group could not be calculated with data reported. **Very low (author-created measure):** Cannot interact with intervention, can only individually predict outcome growth for whole group (not different initial language ability subgroups). Author-created measure of same skill.

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
					Initial language measure and other factors entered into model together. Effect size or direction of effect for each group could not be calculated with data reported.
[13] Smith-Lock et al.	Covariate	Expressive morphosyntax (author created measure)	Expressive morphosyntax (author created measure)	Non-significant relation to outcome growth	**Very low:** Cannot interact with intervention, can only individually predict outcome growth for whole group (not different initial language ability subgroups). Author-created measure of same skill. Effect size or direction of effect for each group could not be calculated with data reported.
[15] Wake et al.	Covariate. Combined model: includes initial language	Age 5 and 6: Expressive vocabulary (expressive	Expressive vocabulary (expressive vocabulary, CELF-P2)	Mean difference is lower when adjusted for, unclear if this is a significant change	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Unclear if change between adjusted

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
	measures with gender and SES [maternal education]	vocabulary, CELF-P2) Word knowledge (recalling sentences, CELF-P2)			and unadjusted analyses is significant. Standardised measures of same and different skill of outcome. Initial language measure and other factors entered into model together. Effect size for each group could not be calculated with data reported.
		Age 6: Expressive vocabulary (expressive vocabulary, CELF-P2) Age 6: Word knowledge (recalling sentences, CELF-P2)	Receptive vocabulary (PPVT-4)	Mean difference is lower when adjusted for, unclear if this is a significant change	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Unclear if change between adjusted and unadjusted analyses is significant. Standardised measure of different skill. Initial language measure and other factors entered into model together. Effect size for each group could not be calculated with data reported.

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
		Age 5 and 6: Expressive vocabulary (expressive vocabulary, CELF-P2) Word knowledge (recalling sentences, CELF-P2)	Word knowledge (recalling sentences, CELF-P2)	Mean difference is lower when adjusted for, unclear if this is a significant change	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Unclear if change between adjusted and unadjusted analyses is significant. Standardised measures of same and different skill of outcome. Initial language measure and other factors entered into model together. Effect size for each group could not be calculated with data reported.
		Age 6: Expressive vocabulary (expressive vocabulary, CELF-P2) Age 6:	Mixed morphosyntax and semantics (Bus Story information, Bus Story subordinate clauses, Bus	Mean difference is lower when adjusted for, unclear if this is a significant change	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Unclear if change between adjusted and unadjusted analyses is significant. Standardised measure of different skill. Initial language

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
		Word knowledge (recalling sentences, CELF-P2)	Story sentence length)		measure and other factors entered into model together. Effect size for each group could not be calculated with data reported.
		Age 5 and 6: Pragmatics (CCC-2) Age 5 and 6: Expressive vocabulary (expressive vocabulary, CELF-P2) Age 5 and 6: Word knowledge (recalling sentences, CELF-P2)	Pragmatics (CCC-2)	Mean difference is higher when adjusted for, unclear if this is a significant change	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Unclear if change between adjusted and unadjusted analyses is significant. Standardised measures of same and different skill of outcome. Initial language measure and other factors entered into model together. Effect size for each group could not be calculated with data reported.

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
		Age 6: Expressive vocabulary (expressive vocabulary, CELF-P2) Age 6: Word knowledge (recalling sentences, CELF-P2)	Phonological awareness (CTOPP)	Mean difference is lower when adjusted for, unclear if this is a significant change	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Unclear if change between adjusted and unadjusted analyses is significant. Standardised measure of different skill. Initial language measure and other factors entered into model together. Effect size for each group could not be calculated with data reported.
Overview of findings and confidence in evidence by outcome	**Expressive vocabulary:** Non-significant correlation between initial language and intervention outcome (1 studies, used different (standardised) measure of same initial language skill and measure of different initial language skill to outcome). Mixed findings for initial language covariates (same and different initial language skills to outcome) predicting intervention outcomes (4 studies). <u>Unclear what the direction or magnitude of effect is when significant [3], and if the lower mean difference in the study [15] from the unadjusted model was significant.</u> **Receptive vocabulary:** Mixed findings for initial language covariates (same and different skills to outcome) predicting intervention outcomes (3 studies). <u>Unclear what the direction or magnitude of effect is when significant [3], and if the lower mean difference in the study [15] from the unadjusted model was significant.</u>				

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
	Word knowledge: Initial language covariates (same and different skills to outcome) seemed to significantly predict/change intervention outcomes (2 studies). Better initial language meant more benefit from the intervention, unclear what the magnitude of effect is for one result [4], and if the lower mean difference in one study [15] from the unadjusted model was significant. **Mixed morphosyntax and semantics:** Initial language covariates (same and different skills to outcome) seemed to change intervention outcome (2 studies). With other variables, seemed to make intervention and control group differences non-significant (only immediately after intervention, but not in follow-up [2]), and unclear if the lower mean difference in the study from the unadjusted model was significant. **Listening comprehension:** Initial language covariates (same and different skills to outcome) appeared mixed in predicting intervention outcome (2 studies). For significant study = better initial language meant more benefit, unclear what the magnitude of effect is. **Expressive morphosyntax:** Mixed findings for initial language covariates (same skills to outcome) predicting intervention outcomes (2 studies). Unclear what the direction or magnitude of effect is when significant for one result [3]. **Pragmatics**: Initial language covariates (same and different skills to outcome) seemed to change intervention outcome (1 study). Unclear if the higher mean difference in the study [15] from the unadjusted model was significant. **Phonological awareness:** Mixed findings for initial language covariates (same and different skills to outcome) predicting intervention outcomes (3 studies, if using the same language skill, was significant; if using general language, was mixed or unclear). If significant =				

Study	'Third variable' analysis method	Initial language measure(s)	Outcome	Result	Confidence in analysis
	better initial phonological awareness meant less benefit [6], and change a non-significant difference between intervention and controls to a significant gain for intervention over controls (when combined with other factors) [2]. Better initial general language meant more benefit [6]. Unclear if the lower mean difference in the study [15] from the unadjusted model (using word knowledge and expressive vocabulary) was significant. The magnitude of effect for all findings was unable to be determined. Confidence in analysis for all outcomes is very low. Confidence based on study numbers for all outcomes is very low.				

Non-verbal IQ

NVIQ was examined in five studies [3, 8,13,15,17] for four outcomes. Analyses suggested there was no moderation effect of NVIQ for expressive morphosyntax and word knowledge outcomes. The confidence in evidence based on number of studies was very low as they were based on a single study [15], and confidence in the analyses was also very low. Findings and confidence judgements are presented in table 2.10.

NVIQ significantly predicted expressive and receptive language growth, but it was unclear what direction of effect this was [3]. NVIQ was also found to be significantly predict and be related to general language (by correlation and covariate analyses [8]), and findings suggested that a higher NVIQ score predicts better general language growth. However, the correlation analyses in the same study found a significant relationship only for the control group rather than the intervention group. NVIQ was found not to relate to [13,17], or predict [3] expressive morphosyntax outcomes. Confidence for evidence based on the number of studies was very low for all outcomes, with evidence being based on a maximum of between one and two studies depending on outcome and analysis type. Confidence in the analyses were very low for all analyses. Findings and confidence judgements are presented in table 2.11.

Table 2.10. Overview of 'third variable' analyses exploring whether different levels in NVIQ cause differential intervention response by study

Study	'Third variable' analysis method	NVIQ measure	Outcome	Result	Confidence in analysis
[15] Wake et al.	Moderation	Unclear, age 5 only Labelled as Specific (non-verbal IQ <85) v non-specific (non-verbal IQ >=85) language delay	Expressive vocabulary (expressive vocabulary, CELF-P2)	Non-significant interaction.	**Very low:** Uses moderation analyses. Specified cut-offs, but not measure. Effect sizes or direction of effect for each group could not be calculated with data reported.
			Word knowledge (recalling sentences, CELF-P2)	Non-significant interaction.	**Very low:** Uses moderation analyses. Specified cut-offs, but not measure. Effect sizes or direction of effect for each group could not be calculated with data reported.
Overview of findings and confidence in evidence by outcome	**Expressive vocabulary:** Non-significant moderating effects of NVIQ on intervention outcome (1 study, very low confidence in analysis). **Word knowledge:** Non-significant moderating effects of NVIQ on intervention outcome (1 study, very low confidence in analysis). Confidence in analysis for all outcomes is very low. Confidence based on study numbers for all outcomes is very low.				

Table 2.11. Overview of 'third variable' analyses exploring how NVIQ relates to outcome growth by study

Study	'Third variable' analysis method	NVIQ measure	Outcome	Result	Confidence in analysis
[3] Dockrell et al.	Covariate	BAS-2	Expressive vocabulary (naming vocabulary, BAS-2)	Was a significant covariate.	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different NVIQ subgroups). Standardised measure. Effect size and direction for each group could not be calculated with data reported.
			Receptive vocabulary (verbal comprehension, BAS-2)	Was a significant covariate.	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different NVIQ subgroups). Standardised measure. Effect size and direction for each group could not be calculated with data reported.
			Expressive morphosyntax (sentence repetition, GAP)	Was a non-significant covariate.	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different NVIQ subgroups). Standardised measure. Effect size and direction for each group could not be calculated with data reported.

Study	'Third variable' analysis method	NVIQ measure	Outcome	Result	Confidence in analysis
[8] Motsch & Ulrich	Correlation	K-ABC-G non-verbal scale	General language (AWST-R)	Significantly and positively correlated with gains in outcome for control group. Non-significant (but negative) correlation with gains in outcome for intervention group.	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different NVIQ subgroups). Standardised measure. Effect size and direction for each group could not be calculated with data reported.
	Covariate		General language (AWST-R)	Was a significant independent predictor.	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different NVIQ subgroups). Standardised measure. Effect size for each group could not be calculated with data reported.
[13] Smith-Lock et al.	Correlation	WNV-3	Expressive morphosyntax (author-created)	Was not significantly related to gain score.	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different NVIQ subgroups). Standardised measure. Effect size and direction for each group could not be calculated with data reported.

Study	'Third variable' analysis method	NVIQ measure	Outcome	Result	Confidence in analysis
[17] Yoder et al.	Correlation	LIPS-R	Expressive morphosyntax (IPSyn)	Was not significantly related to gain score.	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different NVIQ subgroups). Standardised measure. Effect size and direction for each group could not be calculated with data reported.
Overview of findings and confidence in evidence by outcome	**General language:** Significant effect of correlation (1 study, very low confidence). Better NVIQ predicts better outcome for children's general language growth. A significant independent covariate (1 study, very low confidence). Better NVIQ predicts better outcome for children's general language growth. **Expressive vocabulary:** NVIQ was a significant covariate, but it was unclear how (1 study, very low confidence). **Receptive vocabulary:** NVIQ was a significant covariate, but it was unclear how (1 study, very low confidence). **Expressive morphosyntax:** Non-significant correlations between NVIQ and intervention outcome (2 studies, very low confidence). Non-significant finding for NVIQ as a covariate (1 study, very low confidence in analysis). Confidence in analysis for all outcomes is very low. Confidence based on study numbers for all outcomes is very low.				

Co-occurring disorders: Speech

Co-occurring speech difficulties were examined in two studies [6,12] for two outcomes. In both analyses, speech was notable in influencing outcomes. There was a positive difference for Smith-Lock et al. [12] in the magnitude of intervention effect (i.e., effect size grew larger) for expressive morphosyntax when children with speech difficulties were removed from the analysis. Justice et al [6] found that having better speech ability predicted better improvement in one of their two measures of phonological awareness. Confidence for evidence based on the number of studies was very low for both outcomes, with evidence being based on one study each. Confidence in the analyses was very low for phonological awareness, and low for expressive morphosyntax. Findings and confidence judgements are presented in table 2.12.

Table 2.12. Overview of 'third variable' analyses exploring whether speech difficulties affect differential intervention response (via subgroup analysis) and outcome growth (via covariate analysis) by study

Study	'Third variable' analysis method	Speech measure	Outcome	Result	Confidence in analysis
[6] Justice et al.	Covariate	GFTA	Phonological awareness (author created measures of rhyme and alliteration)	Rhyme: was a non-significant predictor of change in the outcome. Alliteration: was a significant and positive predictor of change.	**Very low:** Cannot interact with intervention, can only individually predict outcome growth for whole group (not different speech subgroups). Author created outcome. Effect sizes for each group could not be calculated with data reported.
[12] Smith-Lock et al.	Subgroup	GFTA	Expressive morphosyntax (author created)	There was a difference in effect size when children with speech difficulties were removed from the analysis. Analysis with all children: g = 0.55 Analysis without children with speech difficulties: g = 0.71	**Low:** Represents a comparison between full sample and those without speech difficulties. Uses author-created measure for outcome. Subgroup not as robust for establishing interactions as moderation. Removing children could inflate effect sizes. Different effect sizes provided, appears to be that having speech may reduce intervention effect.
Overview of findings and	**Expressive morphosyntax:** Positive difference when removing children with speech difficulties for intervention outcome (1 study, low confidence in analysis). Better speech meant more benefit from the intervention (moderate to large effects for both subgroups).				

confidence in evidence by outcome	**Phonological awareness:** Mixed finding for covariate analyses (1 study, very low confidence in analysis). Significant finding = better initial speech meant better benefit. Confidence in analysis for all outcomes is between low and very low. Confidence based on study numbers for all outcomes is very low.

Co-occurring disorders: Behaviour

Behaviour was examined as a covariate in one study [2] for five outcomes. The analyses compared the basic model (without covariates) findings with an adjusted model with behaviour (in combination with other factors). There were no differences to the basic model for expressive vocabulary, word knowledge, listening comprehension, mixed morphosyntax and semantics, and phonological awareness. Confidence for evidence based on the number of studies was very low for all outcomes, with evidence being based on one study. Confidence in analyses were also very low. Findings and confidence judgements are presented in table 2.13.

Table 2.13. Overview of 'third variable' analyses exploring whether behavioural difficulties affect outcome growth by study

Study	'Third variable' analysis method	Behaviour measure	Outcome	Result	Confidence in analysis
[2] Bowyer-Crane et al.	Covariate Combined model: Model with initial language of the same measure (depending on outcome), age, gender assigned at birth and behaviour together	SDQ total deviance	Expressive vocabulary (picture naming, WPPSI-3)	Significance of model remained unchanged (measured at follow-up period only).	**Very low (all outcomes):** Cannot interact with intervention, can only predict outcome growth for whole group (not different subgroups). Standardised measure. Effect of change in significance based on a combination of variables, rather than able to say the unique effect of initial language. Effect size or direction of effect for each group could not be calculated with data reported.
			Word knowledge (author created)	Significance of model remained unchanged (measured at immediate post-intervention only).	
			Listening comprehension (NARA-2)	Significance of model remained unchanged (measured at immediate post-intervention period only).	
			Mixed morphosyntax and semantics (Bus Story sentence length,	Significance of model remained unchanged for all three outcomes.	

Study	'Third variable' analysis method	Behaviour measure	Outcome	Result	Confidence in analysis
			Bus Story narrative skill, APT)	(Sentence length and APT measured at immediate post-intervention and follow-up periods; narrative skill measured at immediate post-intervention only).	
			Phonological awareness (SIT, PAT, TPA)	Significance of model remained unchanged for all three outcomes. (All three tests measured at immediate post-intervention only).	
Overview of findings and confidence in evidence by outcome	**Expressive vocabulary, word knowledge, mixed morphosyntax and semantics, listening comprehension and phonological awareness:** Non-significant finding for behaviour as a covariate (in combination with other factors) (1 study, very low confidence in analysis). Confidence in analysis for all outcomes is very low. Confidence based on study numbers for all outcomes is very low.				

Co-occurring disorders: Non-Specific difficulties

Non-specific difficulties were examined as a covariate in one study [10] for two outcomes. The covariate analyses found that non-specific difficulties did not significantly predict expressive or receptive vocabulary growth, and did not moderate intervention outcomes for these language skills. The non-specific difficulties were labelled as special educational status, but not specified further. Confidence for evidence based on the number of studies was very low for all outcomes, with evidence being based on one study. Confidence in analyses were also very low. Findings and confidence judgements are presented in table 2.14.

Table 2.14. Overview of 'third variable' analyses exploring whether non-specific difficulties affect differential intervention response (via moderation analysis) and outcome growth (via covariate analysis) by study

Study	'Third variable' analysis method	Non-specific difficulties measure	Outcome	Result	Confidence in analysis
[10] Pollard-Durodola et al.	Covariate	"Special educational status" – not specified further	Expressive vocabulary (EOWPVT, author created measure)	Non-significant relation to outcome growth	**Very low (all measures):** First set of analyses covariate) cannot interact with intervention, can only predict outcome growth for whole group (not different subgroups). Moderation analyses used also, but still problematic for the following reasons (also applicable to covariate analysis). The measurement of non-specific difficulties is very limited (unclear who is in this group). Two of the four outcomes are author-created. Effect sizes or direction of effect for each group could not be calculated with data reported.
			Receptive vocabulary (PPVT-3, author created measure)	Non-significant relation to outcome growth	
	Moderation - using initial language as an		Expressive vocabulary (EOWPVT,	Did not significantly moderate intervention response.	

Study	‘Third variable’ analysis method	Non-specific difficulties measure	Outcome	Result	Confidence in analysis
	interaction term in multi-level model		author created measure) Receptive vocabulary (PPVT-3, author created measure)	Did not significantly moderate intervention response.	
Overview of findings and confidence in evidence by outcome	**Expressive vocabulary and receptive vocabulary:** Non-significant moderating effect of non-specific and intervention outcome (1 study, very low confidence). Non-significant finding for behaviour as a covariate (in combination with other factors) (1 study, very low confidence in analysis). Confidence in analysis for all outcomes is very low. Confidence based on study numbers for all outcomes is very low.				

Age

Age was examined in four studies [6,7,8,10] for five outcomes. Analyses suggested there was no moderation effect of age for expressive morphosyntax and word knowledge outcomes. The confidence in evidence based on number of studies was very low as they were based on a single study [10], and confidence in the analyses was between low and very low depending on the measure used (i.e., standardised or author-created outcome measure). Findings and confidence judgements are presented in table 2.15.

Table 2.15. Overview of 'third variable' analyses exploring whether age causes differential intervention response by study

Study	'Third variable' analysis method	Outcome	Result	Confidence in analysis
[10] Pollard-Durodola et al.	**Moderation** – using age as an interaction term in multi-level model	Expressive vocabulary (EOWPVT, author created measure)	Did not significantly moderate intervention response.	**Low (EOWPVT):** Uses moderation analyses. Uses standardised measures. No detail of any possible subgroups and effect sizes or direction of effect for each group could not be calculated with data reported. **Very low (author created measure):** Uses moderation analyses. Uses author created measure for outcome. No detail of any possible subgroups and Effect sizes or direction of effect for each group could not be calculated with data reported.
		Receptive vocabulary (PPVT-3, author created measure)	Did not significantly moderate intervention response.	**Low (PPVT-3):** Uses moderation analyses of the same skill, and standardised measure. No detail of any possible subgroups and effect sizes or direction of effect for each group could not be calculated with data reported.

Study	'Third variable' analysis method	Outcome	Result	Confidence in analysis
				Very low (author created measure): Uses moderation analyses of the same skill. Uses author created measure. No detail of any possible subgroups and effect sizes or direction of effect for each group could not be calculated with data reported.
Overview of findings and confidence in evidence by outcome	**Expressive vocabulary and receptive vocabulary:** Non-significant moderating effect age on intervention outcomes (1 study, initial language skills were different to outcomes, between low and very low confidence in analysis). Confidence in analysis for all outcomes is between low and very low. Confidence based on study numbers for all outcomes is very low.			

Age significantly predicted phonological awareness in Justice et al. [6], with older children demonstrating more growth. However, age did not significantly predict growth in general language [8], and expressive and receptive vocabulary [10]. Age also did not significantly correlate with expressive morphosyntax outcomes [7]. Confidence for evidence based on the number of studies was very low for all outcomes, with evidence being based on one study for each outcome. Confidence in the analyses were very low for all analyses. Findings and confidence judgements are presented in table 2.16.

Table 2.16. Overview of 'third variable' analyses exploring how age relates to outcome growth by study

Study	'Third variable' analysis method	Outcome	Result	Confidence in analysis
[6] Justice et al.	Covariate	Phonological awareness (author created measures of rhyme and alliteration)	Rhyme and alliteration: Age is a significant and positive predictor of change for both outcomes.	**Very low:** Cannot interact with intervention, can only individually predict outcome growth for whole group (not different speech subgroups). Author created outcomes. Effect sizes for each group could not be calculated with data reported.
[7] Leonard et al.	Correlation	Expressive morphosyntax (author created measure)	Not correlated with outcome gains for intervention or control group for their specific target.	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different subgroups). Has an author-created outcome. Only examined the targeted outcome for each group (so different implications for each correlation). Effect sizes or direction of effect for each group could not be calculated with data reported.
[8] Motsch & Ulrich	Covariate	General language (AWST-R)	Was not a significantly independent predictor.	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group. Standardised measure. Effect size for each group could not be calculated with data reported.

Study	'Third variable' analysis method	Outcome	Result	Confidence in analysis
[10] Pollard-Durodola et al.	Covariate	Expressive vocabulary (EOWPVT, author created measure) Receptive vocabulary (PPVT-3, author created measure)	Non-significant relation to outcome growth. Non-significant relation to outcome growth.	**Very low (all measures):** Cannot interact with intervention, can only predict outcome growth for whole group. Two of the four outcomes are author-created. No detail of any possible subgroups and effect sizes or direction of effect for each group could not be calculated with data reported.
Overview of findings and confidence in evidence by outcome	**General language:** Non-significant finding for age as a covariate (1 study, very low confidence in analysis). **Expressive vocabulary and receptive vocabulary:** Non-significant finding for age as a covariate (1 study, very low confidence in analysis). **Expressive morphosyntax:** Non-significant finding for age as a covariate (1 study, very low confidence in analysis). **Phonological awareness:** Significant finding for age as a positive covariate (in combination with other factors) (1 study, very low confidence in analysis). <u>The older children are, the more growth in outcomes reported.</u> Confidence in analysis for all outcomes is very low. Confidence based on study numbers for all outcomes is very low.			

Gender assigned at birth

Gender assigned at birth were examined in two studies [10,15] for expressive vocabulary and receptive vocabulary [10,15], word knowledge [15], mixed morphosyntax and semantics [15], pragmatics [15] and phonological awareness [15] outcomes. Gender assigned at birth was not a significant moderator for expressive vocabulary and receptive vocabulary outcomes [10]. When combined in a model with other factors [15], this appeared to change the mean difference slightly for all aforementioned language outcomes, but it was unclear if this was significant. Motsch & Ulrich [8] also completed a correlation analysis with gender, but only for the experimental group (non-significant, no statistics reported), so this is not considered with their other findings. Confidence for evidence based on the number of studies was very low for all outcomes, with evidence being based on one study each per analyses. Confidence in the analyses was between low and very low for expressive and receptive vocabulary analyses, and very low for analyses of the other outcomes. Findings and confidence judgements are presented in table 2.17.

Table 2.17. Overview of 'third variable' analyses exploring whether gender assigned at birth affects differential intervention response (via moderation analysis) and outcome growth (via covariate analysis) by study

Study	'Third variable' analysis method	Outcome	Result	Confidence in analysis
10] Pollard-Durodola et al.	Moderation	Expressive vocabulary (EOWPVT, author created measure)	Did not significantly moderate intervention response.	**Low (EOWPVT):** Uses moderation analyses. Uses standardised measures. No detail of any possible subgroups and effect sizes or direction of effect for each group could not be calculated with data reported. **Very low (author created measure):** Uses moderation analyses. Uses author created measure for outcome. No detail of any possible subgroups and Effect sizes or direction of effect for each group could not be calculated with data reported.
		Receptive vocabulary (PPVT-3, author created measure)	Did not significantly moderate intervention response.	**Low (PPVT-3):** Uses moderation analyses. Uses standardised measures. No detail of any possible subgroups and effect sizes or direction of effect for each group could not be calculated with data reported. **Very low (author created measure):** Uses moderation analyses. Uses author created

Study	'Third variable' analysis method	Outcome	Result	Confidence in analysis
				measure for outcome. No detail of any possible subgroups and Effect sizes or direction of effect for each group could not be calculated with data reported.
	Covariate	Expressive vocabulary (EOWPVT, author created measure)	Non-significant relation to outcome growth.	**Very low (all measures):** Cannot interact with intervention, can only predict outcome growth for whole group. Two of the four outcomes are author-created. No detail of any possible subgroups and effect sizes or direction of effect for each group could not be calculated with data reported.
		Receptive vocabulary (PPVT-3, author created measure)	Non-significant relation to outcome growth.	

Study	'Third variable' analysis method	Outcome	Result	Confidence in analysis
[15] Wake et al.	Covariate Combined model: includes initial language measures with gender and SES [maternal education]	Expressive vocabulary (expressive vocabulary, CELF-P2)	Mean difference is lower when adjusted for, unclear if this is a significant change	**Very low (all measures):** Cannot interact with intervention, can only predict outcome growth for whole group (not different initial language ability subgroups). Unclear if change between adjusted and unadjusted analyses is significant. Standardised measure of different skill. Initial language measure and other factors entered into model together. Effect size for each group could not be calculated with data reported.
		Receptive vocabulary (PPVT-4)	Mean difference is lower when adjusted for, unclear if this is a significant change	
		Word knowledge (recalling sentences, CELF-P2)	Mean difference is lower when adjusted for, unclear if this is a significant change	
		Mixed morphosyntax and semantics (Bus Story information, Bus Story subordinate clauses, Bus Story sentence length)	Mean difference is lower when adjusted for, unclear if this is a significant change	

Study	'Third variable' analysis method	Outcome	Result	Confidence in analysis
		Pragmatics (CCC-2)	Mean difference is higher when adjusted for, unclear if this is a significant change	
		Phonological awareness (CTOPP)	Mean difference is lower when adjusted for, unclear if this is a significant change	
Overview of findings and confidence in evidence by outcome	**Expressive vocabulary and receptive vocabulary:** Non-significant moderating effects of gender assigned at birth on intervention outcome (1 study, very low confidence in analysis). Appears to effect mean difference when gender (alongside other factors) are adjusted for, but unclear if the lower mean difference in the study from the unadjusted model was significant (1 study, very low confidence in analysis). **Word knowledge, mixed morphosyntax and semantics, phonological awareness:** Appears to effect mean difference when gender (alongside other factors) are adjusted for, but unclear if the lower mean difference in the study from the unadjusted model was significant (1 study, very low confidence in analysis). **Pragmatics:** Appears to effect mean difference when gender (alongside other factors) are adjusted for, but unclear if the higher mean difference in the study from the unadjusted model was significant (1 study, very low confidence in analysis). Confidence in analysis for all outcomes is between low and very low. Confidence based on study numbers for all outcomes is very low.			

Socio-economic status

Socio-economic status was examined in four studies, with three examining parental [17]/maternal education [1,15] and one examining area deprivation and free/reduced school meal uptake [2]. Analyses for parental/maternal education included expressive vocabulary [1,15], receptive vocabulary [15], word knowledge [15], mixed morphosyntax and semantics [15], pragmatics [15], phonological awareness [15] and expressive morphosyntax [17] outcomes. Analyses for area deprivation and free school meals included expressive vocabulary, word knowledge, listening comprehension, mixed morphosyntax and semantics and phonological awareness outcomes.

Moderation analysis was only completed in one study examining maternal education [15]. Maternal education did not moderate intervention response for expressive and receptive vocabulary, word knowledge, mixed morphosyntax and semantics, pragmatics and phonological awareness at both ages 5 and 6. Confidence for evidence based on the number of studies was very low for all outcomes, with evidence being based on one study. Confidence in the analyses was between low and very low depending on the time of the analyses (i.e., age 5 or 6). Findings and confidence judgements are presented in table 2.18.

Table 2.18. Overview of 'third variable' analyses exploring whether maternal education causes differential intervention response by study

Study	'Third variable' analysis method	Socio-economic status measure	Outcome	Result	Confidence in analysis
[15] Wake et al.	Moderation	Maternal education (Age 5: high school versus did not finish high school; Age 6: more versus less education but not specified cut-offs)	Expressive vocabulary (expressive vocabulary, CELF-P2)	Did not significantly moderate intervention response (age 5 and 6)	**Low (Age 5, all outcomes):** Uses moderation analyses. Cut-offs specified for SES. Effect sizes or direction of effect for each group could not be calculated with data reported. **Very low (Age 6, all outcomes):** Uses moderation analyses. Cut-offs not specified for SES. Effect sizes or direction of effect for each group could not be calculated with data reported.
			Receptive vocabulary (PPVT-4)	Did not significantly moderate intervention response (age 6 only)	
			Word knowledge (recalling sentences, CELF-P2)	Did not significantly moderate intervention response (age 5 and 6)	
			Mixed morphosyntax and semantics (Bus Story information, Bus Story subordinate	Did not significantly moderate intervention response for all measures (age 6 only)	

Study	'Third variable' analysis method	Socio-economic status measure	Outcome	Result	Confidence in analysis
			clauses, Bus Story sentence length)		
			Pragmatics (CCC-2)	Did not significantly moderate intervention response (age 5 and 6)	
			Phonological awareness (CTOPP)	Did not significantly moderate intervention response (age 5 and 6)	
Overview of findings and confidence in evidence by outcome	Expressive vocabulary, receptive vocabulary, word knowledge, mixed morphosyntax and semantic, pragmatics and phonological awareness: Maternal education did not moderate the effect of intervention outcomes (1 study, between low (age 5) and very low (age 6) confidence in analysis). Confidence in analysis for all outcomes is between low and very low. Confidence based on study numbers for all outcomes is very low.				

Parental education did not significantly correlate with expressive morphosyntax outcomes [17], and maternal education did not correlate with expressive vocabulary outcomes [1]. Maternal education (alongside other factors) did appear to influence the mean difference of expressive and receptive vocabulary, word knowledge, mixed morphosyntax and semantics, pragmatics and phonological awareness outcomes to a small degree [15], but it was unclear if this was significant. Confidence for evidence based on the number of studies was very low for all outcomes, with evidence being based on one study for each analysis. Confidence in the analyses was also very low for all analyses. Findings and confidence judgements are presented in table 2.19

Area deprivation/ free school meal uptake was examined as a covariate in one study [2] for five outcomes. The analyses compared the basic model (without covariates) findings with an adjusted model with SES (in combination with other factors; initial language, age, and gender assigned at birth) included. There were no differences to the basic model for expressive vocabulary, word knowledge and listening comprehension growth. For mixed morphosyntax and semantics, results from one of the three measures from showing a significant group difference group (in favour of intervention group) to showing a non-significant difference at immediate post-intervention. However, there was no change in significance at the follow-up period. For phonological awareness, results for one of the three measures used also changed significance, where a non-significant group difference became significant (in favour of intervention group). All three measures were only tested immediately after intervention, so findings for this were only applicable to this time point. What these findings could indicate is unclear, as no effect size or direction of effect was provided. While it affects significance, the combination with other factors and lack of statistical reporting makes its effects hard to parse apart. Confidence for evidence based on the number of studies was very low for all outcomes, with evidence being based on one study. Confidence in the analyses was also very low for all analyses. Findings and confidence judgements are presented in table 2.19.

Table 2.19. Overview of 'third variable' analyses exploring how maternal education, and area deprivation/free school meal uptake relates to outcome growth by study

Study	'Third variable' analysis method	Socio-economic status measure	Outcome	Result	Confidence in analysis
[1] Aguilar et al.	Correlation	Maternal education (in years)	Expressive vocabulary (author created measure)	Non-significant relation to outcome growth	**Very low:** Cannot interact with intervention, can only relate to outcome growth for whole group (not different initial language ability subgroups). Defined maternal education as a continuous variable, but no r value provided to determine direction. Outcome is author created. Effect size or direction of effect for each group could not be calculated with data reported.
[2] Bowyer-Crane et al.	Covariate Combined model: Model with initial language of the same measure	Area deprivation score and child in receipt of free school meal uptake (government measures) – unclear if both of these were	Expressive vocabulary (picture naming, WPPSI-3) Word knowledge (author created)	Significance of model remained unchanged (measured at follow-up period only). Significance of model remained unchanged (measured at immediate post-intervention only).	**Very low (all outcomes):** Cannot interact with intervention, can only predict outcome growth for whole group (not different subgroups). Standardised measure for outcome, and SES based on government measures, but unclear if both of

Study	'Third variable' analysis method	Socio-economic status measure	Outcome	Result	Confidence in analysis
	(depending on outcome), age, gender assigned at birth and SES together)	used/how these were used as a covariate	Listening comprehension (NARA-2) Mixed morphosyntax and semantics (Bus Story sentence length, Bus Story narrative skill, APT)	Significance of model remained unchanged (measured at immediate post-intervention period only). Bus Story sentence length: No change to significance of initial group comparisons (measured at immediate post-intervention and follow-up periods). Bus Story narrative skill: No change to significance of initial group comparisons (measured at immediate post-intervention period only). APT: Changed significant difference between intervention and control groups (in favour of intervention group) to non-significant at immediate post-	these were used/how these were used as a covariate. Effect of change in significance based on a combination of variables, rather than able to say the unique effect of initial language. Effect size or direction of effect for each group could not be calculated with data reported.

Study	'Third variable' analysis method	Socio-economic status measure	Outcome	Result	Confidence in analysis
				intervention; no change in significance at follow-up period.	
			Phonological awareness (SIT,PAT,TPA)	SIT: Changed non-significant difference between intervention and control groups to significant (in favour of intervention group, measured at immediate post-intervention period only). PAT and TPA: Significance of model remained unchanged. (both tests measured at immediate post-intervention only).	
[15] Wake et al	Covariate Combined model: initial language, gender assigned at birth with SES)	Maternal education (Age 5: high school versus did not finish high school; Age 6: more versus less education but not specified cut-offs)	Expressive vocabulary (expressive vocabulary, CELF-P2)	Mean difference is lower when adjusted for, unclear if this is a significant change	**Very low (all outcomes):** Cannot interact with intervention, can only predict outcome growth for whole group (not different SES subgroups). Unclear if change between adjusted and unadjusted analyses is significant. Standardised measures. SES and other factors
			Receptive vocabulary (PPVT-4)	Mean difference is lower when adjusted for, unclear if this is a significant change	

Study	'Third variable' analysis method	Socio-economic status measure	Outcome	Result	Confidence in analysis
			Word knowledge (recalling sentences, CELF-P2)	Mean difference is lower when adjusted for, unclear if this is a significant change	entered into model together. Effect size for each group could not be calculated with data reported.
			Mixed morphosyntax and semantics (Bus Story information, Bus Story subordinate clauses, Bus Story sentence length)	Mean difference is lower when adjusted for, unclear if this is a significant change	
			Pragmatics (CCC-2)	Mean difference is higher when adjusted for, unclear if this is a significant change	
			Phonological awareness (CTOPP)	Mean difference is lower when adjusted for, unclear if this is a significant change	

Study	'Third variable' analysis method	Socio-economic status measure	Outcome	Result	Confidence in analysis
[17] Yoder et al.	Correlation	Parental education (Nine point scale of years in school with 7 = at least 4 years in university)	Expressive morphosyntax (IPSyn)	Was not significantly related to gain score.	**Very low:** Cannot interact with intervention, can only predict outcome growth for whole group (not different SES subgroups). Standardised measures. Effect size and direction for each group could not be calculated with data reported.
Overview of findings and confidence in evidence by outcome	Area deprivation and reduced/free school meal uptake: **Expressive vocabulary, word knowledge and listening comprehension:** Did not significantly change outcomes when accounted for (1 study, very low confidence in analyses). **Mixed morphosyntax and semantics:** Mixed findings for influencing model significance when SES (in combination with other factors) are accounted for (1 study, very low confidence in analyses). Significant result = depending on SES (and other factors) children could be gaining less benefit from intervention (but only immediately after intervention and not longer term (as follow-up did not change)). **Phonological awareness:** Mixed findings for influencing model significance when SES (in combination of other factors) are accounted for (1 study, very low confidence in analyses). Significant result = depending on SES (and other factors) children could be gaining more benefit from intervention (immediately after intervention). Parental/Maternal education:				

Study	'Third variable' analysis method	Socio-economic status measure	Outcome	Result	Confidence in analysis
	Expressive vocabulary: Maternal education did not significantly correlate with outcomes (1 study, very low confidence in analysis). Appears to effect mean difference when maternal education (alongside other factors) are adjusted for, but unclear if the lower mean difference in the study from the unadjusted model was significant (1 study, very low confidence in analysis). **Receptive vocabulary, word knowledge, mixed morphosyntax and semantics, and phonological awareness:** Appears to effect mean difference when maternal education (alongside other factors) are adjusted for, but unclear if the lower mean difference in the study from the unadjusted model was significant (1 study, very low confidence in analysis). **Expressive morphosyntax:** Parental education did not significantly correlate with outcomes (1 study, very low confidence in analysis). **Pragmatics:** Appears to effect mean difference when maternal education (alongside other factors) are adjusted for, but unclear if the higher mean difference in the study from the unadjusted model was significant (1 study, very low confidence in analysis). Confidence in analysis for all outcomes is very low. Confidence based on study numbers for all outcomes is very low.				

2.3.9. Reporting bias

Publication bias

A qualitative assessment of publication bias will be presented to explore potential influential issues. Three areas of publication bias are explored via: 1) language and country of publications, 2) positive publications (i.e., reporting non-significant effects), and 3) references (author overlap). Missing data is also assessed. There will also be further interpretation as to what that individual risk of bias could indicate when studies are considered together.

Language and country of publications

Bias to non-English speaking interventions and the inclusion of a limited range of countries may have been introduced due to only having the resources to utilise publications in English. Specifically, there was an overrepresentation of English-speaking countries and interventions within the included studies. Around half of the intervention studies were from the USA, followed by the UK (4), Australia (3 studies, but 4 papers) and Germany (1). There is no representation of non-English interventions except for Motsch and Ulrich [8]. Although the UK, USA and Australia are not the only countries with English as one of (or solely) their national languages, these countries are generally regarded as heavily westernised cultures. However, it would be misleading to assume that similar backgrounds apply across these countries in terms of theoretical paradigms, measurement, funding source and amount, procedure and practice. For example, assuming the UK, US and Australia would be similar because they are all English-speaking countries would not be appropriate. Differences have been found in in how they qualify speech and language practitioners, fund intervention, and provide intervention to children with DLD (e.g., Forsythe et al., 2020; McKean et al., 2019). Australia and the US also have largely dissimilar health systems to the UK, Germany, and each other (e.g., Glover & Woods, 2020; The Commonwealth Fund, 2020), which will have differential influences on their speech and language intervention practices. Of note also was that when examining study types used in countries, the US conducted exclusively QE studies, while the majority of the UK, Australian and German studies were RCTs (with the exception of Dockrell et al. [3] and Smith-Lock et al. [12] from the UK and

Australia respectively). Therefore, findings should only be interpreted in the context of predominantly English-speaking, western interventions, bearing in mind certain study types may be more common in particular countries.

Positive publication bias: Reporting of non-significant findings

Both significant and non-significant results (both for overall intervention efficacy and 'third variable' analyses) were reported in each study. The only cases in which there was a possibility for positive publication bias was in the papers where explicit significance levels for 'third variable' analyses were not given [5, 14, 16]. Therefore, there appeared to be little evidence to suggest that positive publication bias was evident to a large extent.

Potential reference bias: Author overlap

Two papers have the same authors, with one prior study appearing to be a trial for the larger scale intervention later [12, 13]. But although the authors are the same in both Smith-Lock et al. papers [12, 13] and have some similarities in the intervention they used, the scale, sample and some procedures did differ. In addition, the 'third variable' analyses differed between the two studies ([12] was comorbidity, [13] was initial language and NVIQ). As such, their impact on the validity of the overall synthesis is minimal and unlikely to introduce bias. The Bowyer-Crane [2] and Haley et al. [5] papers shared some (but not all) authors, which may have introduced bias via overlap, but the studies utilised different samples, interventions, 'third variable' analyses and methodologies. This would indicate that the studies were different in nature from and effects would not be likely subject to bias. The rest of the studies were by different research teams and authors.

Missing data

In three studies [2,5,15], some participants were missing in the final analysis, and it was unclear how many were missing (because ranges rather than exact numbers for analyses were provided). For two studies [2,5], this missing data was reported as due to attendance and co-operation issues. No reason was provided in Wake et al. [15]. As mentioned in selective reporting sections, some outcome and 'third variable' analyses data was missing. Authors were contacted to provide missing data, but either no response was received, or data was not available. As data synthesis was narrative with the support of quantitative data, it was not possible to perform sensitivity analysis or data imputing as suggested by Cochrane (Higgins et al., 2021).

2.3.10. Certainty of evidence

The GRADE criteria were used to determine the certainty of evidence. Two out of the five domains (risk of bias, imprecision) were rated very low, and two aspects of the evidence (inconsistency, publication bias) were rated low. Only indirectness was rated with high confidence. Therefore, there is an overall low confidence in the synthesised evidence. Reasonings for confidence ratings by each domain are provided in table 2.20.

Table 2.20. GRADE certainty ratings and reasons

GRADE domains	Certainty rating	Reasons
Risk of bias	Very low confidence	Varied risk of bias across studies, and all studies had some level of high or unclear risk.
Imprecision	Very low confidence	Absolute effects could not be estimated from 'third variable' analyses due to lack of statistical data. Confidence for almost all 'third variable' analyses were rated either low or very low (the only exception being analyses for language profile, which were rated as moderate).
Inconsistency	Low confidence	A small number of studies represented each finding, and results (even for similar 'third variable' analyses types) are mixed.
Indirectness	High confidence	All studies used interventions in the populations of interest (e.g., using interventions which would be used in actual practice for children with language difficulties), and studied realistic outcomes (measures of oral language skills).
Publication bias	Low confidence	There is evidence of bias in two of the four assessments of publication bias (language and country, positive results).

2.4. Discussion

The systematic review and narrative synthesis examined which child- and social factors are associated with differential responses to preschool language interventions for children with language difficulties. To answer the research question, two objectives were addressed; 1) "What participant factors are described in intervention studies for preschool language, and have been included in 'third variable' analyses, and why?", and 2) "What conclusions can be drawn regarding the impact of the identified child and social factors on preschool language intervention response?". A discussion of the findings extracted to answer each objective, strengths and weaknesses of the study, and overall conclusion are presented in the following section.

2.4.1. Overall findings

What participant factors are described in intervention studies for preschool language, and have been included in 'third variable' analyses, and why?

Each study commonly described the chosen factors for their samples. The information about factors (i.e., measures, thresholds, subgroups within the factor) were usually reported in some detail. Therefore, what was reported would generally give an idea of factors of participants receiving interventions, and this appeared to be important to all researchers. But factors were inconsistently and rarely subject to 'third variable' analyses, with theoretical or statistical justification being seldom reported in most studies. Many of the studies completed group comparisons for factors to ensure group equivalency, and some chose to continue with, or not complete 'third variable' analyses based on this. This meant that it was difficult to carry out my analysis as many potential factors were not examined. While group comparisons could indicate that differences in outcomes were not due to differences between groups, it does not allow for an understanding whether differential intervention outcomes occur based on different levels of the factor. It is not encouraged to simply analyse all possible factors, as this would increase the possibility of 'false positive' results. However, there should be *a priori* consideration on how some factors could be theoretically important to impacting change, rather than just considering differences at the start of the intervention. As such, future studies which examine child and social factors' effects on intervention response should still complete 'third variable' analyses with factors, even if groups are equivalent.

As noted previously in the introduction chapter, and within this chapter (selective reporting of third variable analyses, section 2.3.6), intervention studies to date have not been set up, or do not prioritise analyses considering child and social factors, and are more focused on determining if interventions work. While determining whether interventions work generally is important to examine initially, it is clear some researchers acknowledge that factors could influence intervention response. However, more work needs clearly needs to be done to expand our knowledge on intervention efficacy by changing the question from 'does this intervention work?' to 'does this intervention work, and if so, for whom?'. 'Third variable' analyses for some factors may be difficult to analyse for practice or policy reasons. For example, some studies had samples focusing on a specific subgroup (e.g., children from low SES backgrounds). Prior literature and recommendations from policy for many years suggests focusing on specific groups with special educational needs like language difficulties is important, as they may require more or focused intervention (e.g., Lindsay et al., 2020; Select Committee on Education and Skills, 2006). As such, in some research contexts including other subgroups of that factor are considered inappropriate. For example, children from low SES backgrounds may have specific challenges that also need addressing which children from middle and upper SES backgrounds may not (e.g., low educational resources at home). In intervention studies like this, it would not make sense to analyse a single-level factor. When examining the synthesised data, studies with a focus on specific subgroups did not explore all of their other reported factors. This is an issue because it is still important to consider how benefit from an intervention may be different for a targeted subgroup when other risk- or protective factors are looked at. For example, children from low SES backgrounds are likely to be exposed to a number of risk factors (Greenwood et al., 2020) that may influence intervention. It is also well acknowledged that children requiring language intervention are heterogeneous in multiple areas (Bishop et al., 2017). As noted in the introduction chapter, understanding the combined factors and risks for children is essential in understanding if and how differential intervention effects occur for these children in language interventions, which will help construct better interventions to compensate for such potentially combined risks. This will potentially benefit at risk groups more than current practice. As such, more work is needed in intervention research in choosing ‘third variable’ analyses for described factors. It is recommended that future studies need to make explicit hypotheses about, and include, reported factors in ‘third variable’ analyses (or justify why not). This is especially important as it will help determine how these factors relate to intervention response, and potentially how they work together to compound or alleviate differences in interventions between children.

What conclusions can be drawn regarding the impact of the identified child and social factors on preschool language intervention response?

Findings suggested that differences in initial language (of the same and different skill) and speech affected both outcome growth and intervention response. Findings also suggested that NVIQ, age and SES (area deprivation and free school meal uptake) relate to language outcome growth.

Overall, initial language appeared to relate to language growth for almost all language skills. In most cases, language growth related to initial language which was the same skill as the outcome (e.g., initial vocabulary relating to vocabulary growth). Findings were more mixed for initial language skills different to the outcome (e.g., general language did not predict alliteration outcomes [6]; initial receptive vocabulary, word knowledge and general language did predict listening comprehension outcomes [4]) and if a different measure was used (e.g., initial EVT-2 did not relate to taught vocabulary learned [1], mean length of utterances did relate to productive syntax [17]). In most cases however, it was unclear what the magnitude and direction of these relationships were, with the exception of a few studies highlighted below.

The significant findings for initial language ability indicated having more severe initial general language, word knowledge, and expressive morphosyntax difficulties meant children gained more from language intervention, and children with milder difficulties benefitted less. In all cases, these initial abilities were examined for the same skill measured for the outcome, except for initial general language which affected both general language [11] and phonological awareness [6] outcomes. This is counter to the hypothesis that children would gain less from interventions as their weaker language skills would make it harder for them to understand and engage in some steps/tasks related to those skills in the intervention (Storkel et al., 2017). This is good news, as this indicates children with more severe language difficulties are able to catch up if difficulties are targeted. Why children with milder difficulties gained less may likely be due to having less gains to make. However, children with more severe difficulties gaining more from interventions may also not be a true effect, due to regression to the mean. Specifically, if there is random error from assessments, then initially low scorers are more likely to increase than decrease their score after an intervention (Linden, 2013). While random error is smaller in RCT studies and so this is unlikely to be an issue (i.e., for the general language result [11]), the findings for word knowledge and expressive morphosyntax were based on quasi-experimental studies where results are more subject to random error. As such, it may be that results for word knowledge and expressive morphosyntax are not a differentiating effect and instead due to a regression to the mean.

In comparison to the other oral language skills, initial listening comprehension was shown to have a mixed pattern of effect for listening comprehension outcomes. Children with milder language difficulties gained more from intervention than children with more severe difficulties in a study using covariate analysis [4], but gained less in a study using moderation analysis. To explain these findings, it is important to consider the underlying mechanisms for listening comprehension [9]. Why this could be is because listening comprehension is underpinned by many complex cognitive processing abilities, and also draws upon a wide range of language skills including syntactic and vocabulary skills (Kim & Pilcher, 2016). Therefore, children with higher listening comprehension may have an advantage because it may be harder for children with poorer listening comprehension to catch up if they also score poorly in these other language areas. This may explain the findings of the covariate study, because the analysis examined a combination of good initial listening comprehension and other language skills (e.g., receptive vocabulary, general language, word knowledge) to predict outcomes, and focused on treating these skills in the intervention. This did not explain the moderation study findings, as they also had tasks which targeted language skills associated with listening comprehension (although they were not included in the same analysis). But it is suspected that the moderation findings may have also been subject to regression to the mean like findings for word knowledge and expressive morphosyntax due to being a quasi-experimental study. It should be noted however that these findings are represented by two studies, and so more research needs to be done to understand how children's listening comprehension intervention gains differ based on their initial listening comprehension difficulties.

When considering these difficulties together, it was found that the severity of initial language difficulties may differ in direction of effect depending on the oral language skill. These differences may be based on the complexity of underlying mechanisms that support development of the language skill (e.g., Kim & Pilcher, 2016), but it is unclear if this is the case from the available evidence. This supports research that some receptive language skills may be harder to treat than expressive language (Boyle et al., 2007; Boyle et al., 2010). However, where different effect sizes were reported (general language, listening comprehension, word knowledge, expressive morphosyntax) the magnitude of intervention effect for all children was still between small and moderate. This means that while children do have differing intervention response based on their initial language, they still gained from language interventions. This is encouraging, as it appears that language intervention is suitable for treating children with different levels of difficulty, and can also target both receptive and expressive difficulties. Therefore, these findings also support literature that language comprehension can be treated effectively (Broomfield & Dodd, 2011; Ebbels et al., 2014, 2017). While both positions may at first appear to contradict one-another, it may be that it is difficult to treat receptive skills unless other weaker language and cognitive abilities associated with their development are also considered and treated.

There were non-significant moderation findings for initial language for expressive vocabulary, receptive vocabulary, mixed morphosyntax and semantics, pragmatics, and phonological awareness outcomes. This is potentially good news, as children with differing levels of difficulty may benefit equally in interventions addressing these outcomes. This also indicates that the severity of initial language difficulties may be more important for some oral language skills, and not for others. There was also a non-significant finding for language profile (expressive/ receptive/ mixed) for both expressive vocabulary and word knowledge outcomes. This could indicate children with different language difficulty profiles gain equally from intervention, and again runs counter to Boyle et al. (2007; 2010) that children with receptive difficulties are harder to treat. But as discussed in the strengths and limitations section, the findings for these studies are based on a small number of studies, and/or the quality of their analyses are generally low. In addition, intervention effects in studies examining these initial language skills and language profile as moderators were non-significant, and so it may be difficult to determine if differential intervention response occurred based on initial language severity or language profile if the intervention did not benefit the children. Therefore, it will be important to examine initial language as a moderator for intervention effects which are significant to verify these findings.

Speech difficulties (specifically articulation) also affected intervention response for expressive morphosyntax and growth for phonological awareness. Specifically, children with better speech skills benefitted more in their expressive morphosyntax and phonological awareness from language interventions, and those with worse speech benefitted less. This supported the hypothesis posited that because weaker speech undermines oral language development (Haskill & Tyler, 2007; Lewis et al., 2015), this could also potentially undermine the extent of children's gains in their language via intervention. However, speech was only examined in interventions with language skills which are strongly related to speech (Dodd et al., 2018; Murray et al., 2019). For example, children may know their third person singular –{s}, but their speech prevents them from being able to pronounce it. Regardless, it is important to address speech difficulties in language interventions if they are present. On one hand, if speech is found in future research to produce differential intervention responses in less related oral language skills, then this should be addressed. If poor speech is incorrectly conflated with the child's language knowledge with their ability to signal phonemes or morphemes expressively, this will prevent a clear understanding of intervention efficacy. Therefore, more research needs to be completed to determine which of these is the case.

In the instance an effect size was reported (expressive morphosyntax, [12]) the magnitude of intervention effect before removing children with speech difficulties was still moderate. This means that while children with speech difficulties benefitted less and may potentially require more intensive intervention, they still gained in expressive morphosyntax. It is therefore likely important to include children with additional speech difficulties in language interventions.

Relationships were demonstrated between NVIQ, age and SES (area deprivation/free school meal uptake) and language growth. Area deprivation/free school meal uptake appeared to change the significance for mixed morphosyntax and semantics and phonological awareness. However, its specific effects on these outcomes were unclear. As such, it could not be determined if findings supported the hypothesis made that because language interventions can address the deficits in resources promoting language at home, children living in social disadvantage could benefit more from language interventions (McKean et al., 2015, 2017). Furthermore, it is difficult to ascertain whether area deprivation/free school meal uptake related to intervention efficacy or not for both outcomes due to the type and quality of the analyses. So, while the effects are unclear, area deprivation/free school meal uptake could be an important moderator as it appears to have an effect on language growth.

In line with prior literature (Ebert, 2021; Griffiths et al., 2022; Smolak et al., 2020; Snijders et al., 2020; Willinger et al., 2019; Yim & Yang, 2018), better non-verbal IQ appeared to positively predict general language, expressive and receptive vocabulary gains form interventions. However, NVIQ did not predict expressive morphosyntax gains. NVIQ has been shown to have an inconsistent relationship with morphosyntax, and is considered to develop independently from cognitive abilities (Dethorne & Watkins, 2006). In contrast, vocabulary appears to have a reciprocal relationship with NVIQ (Griffiths et al., 2022), which demonstrates why NVIQ associated with expressive and receptive vocabulary growth. NVIQ may have also associated with general language growth because the measure examines vocabulary learning (AWST-R, [8]).

Furthermore, NVIQ was also a non-significant moderator for expressive vocabulary and word knowledge intervention gains. This finding is counter to the hypothesis made that children scoring lower on NVIQ assessments have more general cognitive difficulties which provide a barrier to their engagement with learning activities (Alibali & Nathan, 2018), and so gain less because they find it difficult to engage with learning tasks in interventions. Instead, findings support the previous research seen for older children, that NVIQ does not impact language intervention response (Boyle et al., 2007). Why this result occurred could be due to the intervention itself. For example, Boyle et al. (2007) examined interventions implemented by SLTs and SLT assistants. Speech and language practitioners tend to employ child-centred approaches for their interventions based on the specific needs of their children (Forsythe et al., 2020). As such, it may be that NVIQ was not a moderator in Boyle because of SLTs and assistants would have been likely to take into account children’s cognitive difficulties. This is also seen in the study with the non-significant moderator result [15], as they explicitly state “we designed a program that is both standardized and replicable, yet flexible enough for children with diverse cognitive and language profiles” (p897, Wake et al., 2013). With this in mind, children may have had similar gains from the language intervention because their cognitive differences were accounted for by the intervention. However, this result may instead be because the intervention effects for expressive vocabulary and word knowledge were non-significant. Therefore, it may be difficult to determine if differential intervention response occurred by NVIQ subgroups if the intervention did not benefit children. However, this finding is based on only one study, so more research examining the effects of NVIQ are needed, especially comparing interventions which do (e.g., child-centred) and do not (e.g., more prescribed interventions) explicitly account for children’s cognitive differences.

Analyses for age yielded mostly uninterpretable or non-significant findings for general language, expressive vocabulary and receptive vocabulary. However, a single significant finding in Justice et al. [6] suggested being older meant better growth in phonological awareness (rhyme and alliteration skills). This supports the hypothesis that older children may benefit more in interventions because they are more experienced in educational tasks; and they are generally more cognitively and socially developed which allows them to access learning more easily (Cantalini-Williams et al., 2016). This also supported evidence that rather than performing poorly after a ‘critical period’ of language development, children older than 5 can change and still benefit significantly from language intervention (McKean et al., 2015; Taylor et al., 2013). This is important as it suggests early intervention may not always be the most successful method for improving some language difficulties that children have. As such, interventions should still be being implemented with older children with phonological awareness difficulties as they still can change and benefit significantly from language intervention. Why age was significant only for phonological awareness may be because unlike the other language skills examined, phonological awareness is a metalinguistic skill that requires a high level of executive control to develop well (Friesen & Bialystok, 2012; Gombert, 1997). As such, it may be that executive control requires children to be more developmentally mature to access further learning for this skill. Similar to findings for the other factors, more studies are required to verify these effects findings reported here are based on a small number of studies.

Finally, behaviour, maternal education, gender assigned at birth and non-specific difficulties were not clearly or did not significantly moderate intervention response or outcome growth. The findings for maternal education support the hypothesis that children from lower SES backgrounds and with low language may be able to catch up if language is monitored and targeted by intervention, and if their home learning and literacy environment is optimal (McKean et al., 2015, 2017). Furthermore, while behaviour has been shown to be linked to preschool oral language development (Vermeij et al., 2021), it may support the hypothesis that co-occurring behavioural difficulties may not have internal mechanisms that impact language intervention response. The moderation findings for gender assigned at birth were unexpected based on Boyle et al. (2007) who found females benefitted more than males in language intervention outcomes. These findings may be encouraging, as they indicate that male and female children, children from different SES backgrounds, and children with and without behavioral difficulties may benefit equally in interventions. But as seen with other factors, intervention effects were non-significant in studies where these factors were analysed. Therefore, it may be difficult to determine if differential intervention response occurred as these interventions did not benefit children. Non-specific difficulties were defined very vaguely in included studies and so it would be unclear what to conclude from findings even if significant. Overall, like the other factors, behaviour, maternal education, and gender assigned at birth require more high-quality moderator research to be able to draw appropriate conclusions.

2.4.2. Strengths and limitations

The systematic review to the author's knowledge was the first to conduct a systematic and comprehensive examination of the current available evidence of analyses exploring how a number of child and social factors may produce differential language intervention response for preschool-aged children. It was able to highlight what factors of interest were reported in samples, and had an inclusive approach to analyses used by researchers when examining the effects of child and social factors on intervention response. Furthermore, it was able to identify research gaps in conducting and reporting such analyses.

However, there were limitations concerning the availability and quality of the evidence. As seen when examining risk of bias for individual studies and overall bias (according to GRADE and publication bias) it was likely that the findings were subject to a high degree of bias, and there was little confidence in the evidence. Studies were typically rated with unclear or high risk of bias in most categories, and findings were predominantly from quasi-experimental studies, and based in English-speaking and western cultured countries. Furthermore, assessment of bias indicated there was selective reporting for ‘third variable’ analyses for most studies. Confidence in analyses were generally rated low and very low. There were at most four studies for each factor per outcome, and this was only for some outcomes and analyses of initial language ability. For all other factors, findings were commonly based on one or two studies, and produced inconsistent findings. Further, Hedges *g* effect sizes or the direction of effect for ‘third variable’ analyses were impossible to acquire in most cases due to the lack of statistical data reported. In addition to bias, a number of the results (i.e., the covariate and correlation analyses) did not reflect a direct association with the intervention, but instead could only estimate how it individually predicted or related to outcome growth. Only a handful of studies included analyses which could determine interactions (moderation), and no mediation analyses were found in papers. It was difficult to group studies together due to how different they were, which meant meta-analyses could not be conducted. Finally, a number of ‘third variable’ analyses were uninterpretable, which reduced the ability to draw further findings from the data. Taken together, these issues with the data meant findings were highly tentative and incomplete.

Although it is disappointing not to understand the extent of these factors fully, the systematic review was essential in understanding the lack of information that is currently present. Specifically, these findings highlight the importance for explicitly reporting analyses choices, statistics and implications. If further research is completed to build on these findings with this in mind, it will lead to better quality results and robust findings. This will then lead to stronger evidence on how to provide effective allocation of support and help interventionists better target children’s needs for language development and in turn school readiness. The current study drew the most out of what was available. But to understand if and how factors affect preschool language intervention response, researchers and other practitioners need to consider how the child- and social factors of their samples explicitly interact with them.

Furthermore, as part of the current review, the aim was to also examine ACEs and multilingualism. However, without a body of evidence that can robustly or even tentatively support the relation between ACEs or multilingualism and language development (while the other factors have this), it would not feel appropriate to explore these in the current thesis. Like many studies before it, ACEs and multilingualism status have been excluded as they are considered too complicated to examine, particularly because there is not enough data to establish good analyses of them. To stop this from re-occurring in research, a large shift in how researchers and interventionists conduct their methods and discovery science is desperately needed.

The review had an inclusive approach to any outcome measure, analysis conducted for ‘third variable’ analyses due to anticipating there would be little data available. Generally in systematic reviews, it is advised that one outcome/ measure is collected per study (McKenzie et al., 2021). This is because effects may not be independent of each other due to the analysis utilising the same participants (McKenzie et al., 2021). However, removing additional measures and focusing on one outcome would mean significantly less findings from an already limited pool of information. Being too restrictive with data can introduce bias (as selection of some results inevitably results in actively not reporting other results), and limit the theoretical understanding of a topic (Heesen et al., 2018). Therefore, once theory and hypotheses are more established (which has begun based on the findings included in this review); and there is a bigger pool of studies in the future, it may be more appropriate to be restrictive in what is analysed. As such, the current review was able to provide a good understanding of the current evidence in the field which can now be built upon in future research.

2.4.3. Conclusions

Overall, participant factors are generally described well in studies, but if, how and why they are chosen as ‘third variable’ analyses are relatively unclear and limited. However, the findings also indicate that researchers are somewhat acknowledging that certain attributes of their participants are important for the context of the intervention. Different abilities in initial language and speech were related to different intervention response. Though, this may be dependent on the oral language dimension treated, and children generally still benefit from interventions despite their differences in these areas. Age and SES predict language outcomes, but the current study was unable to find analyses which examine their differential impact on language intervention outcomes. NVIQ may potentially also be a factor worth examining, but while evidence suggests that this may be important for general language growth, it may be less important if children are receiving an intervention. Language profile, behaviour, gender assigned at birth and non-specific difficulties were non-significant in analyses. The findings presented are likely prone to bias, and the confidence in evidence was judged to be low or very low in almost all areas. The majority of findings were based on predicting or relating to intervention growth, rather than how different levels of a factor affected intervention response. As such, the current evidence is tentative and very limited. Much more work is needed to directly address these questions. Better reporting and changes in methodological approach to intervention research is required to fully address whether different child and social factors relate to differential intervention response.

2.5. Chapter 2 appendices

2.5.1. Appendix A. Review registering and checking for review duplications via PROSPERO

The review protocol was submitted to PROSPERO (registration number: CRD42019119934), a systematic review submission database for topics with health-related outcomes (https://www.crd.york.ac.uk/prospero/). PROSPERO was chosen as it is widely used and supported by the review community, with over 30,000 registered reviews (Page et al., 2018). It also has a database of its registered reviews that are both completed and ongoing. Many of the reviews are also registered in other review databases (e.g., Cochrane). This is useful to prevent unintended duplication of other reviews. Furthermore, submitting to PROSPERO ensured transparency during the review process (e.g., any changes to the original review plan are documented) which helped the review maintain a high-quality standard.

The present review was compared against registered reviews in the PROSPERO database by using search terms of the integrated MESH index to check for similar reviews. The main attributes looked for were if the participants were in the relevant age group and language difficulty; and if the review examined the effect of child and social factors on intervention response. As of December 2018, there were no comparable reviews in these areas. Review protocols found and search terms used are listed in tables 2.21 and 2.22.

Table 2.21. Stage 1: The keyword checking process results within MeSH.

Keyword	**MeSH Terms found when searching keyword in PROSPERO [under 'Stem']**	**Definition of Keyword under MeSH**
Child [under 'Stem' index]	Child; Child Behaviour; Child Behaviour Disorders; Child Development; Child Development disorders, Pervasive; Child Health; Child Language; Child, Preschool	A person 6 to 12 years of age. An individual 2 to 5 years old is CHILD, PRESCHOOL.
Child, Preschool [under 'Permute' index] Changed from 'Preschool' to 'Child, Preschool'	Child, Preschool; Children, Preschool; Preschool Child; Preschool Children	A child between the ages of 2 and 5.
Developmental Language Disorder [under 'Permute' index]	Language Development Disorders; Developmental Language Disorder; Developmental Language Disorders	[Language Development Disorders definition] Conditions characterized by language abilities (comprehension and expression of speech and writing) that are below the expected level for a given age, generally in the absence of an intellectual impairment. These conditions may be associated with DEAFNESS; BRAIN DISEASES; MENTAL DISORDERS; or environmental factors.
Oral Language [under 'Stem' and 'Permute' indexes]	No MeSH term	No MeSH term definition

Keyword	MeSH Terms found when searching keyword in PROSPERO [under 'Stem']	Definition of Keyword under MeSH
Expressive language [under 'Stem' and 'Permute' indexes]	No MeSH term	No MeSH term definition
Receptive language [under 'Stem' and 'Permute' indexes]	No MeSH term	No MeSH term definition
Vocabulary [under 'Permute' index]	Language tests; Vocabulary, Vocabulary Controlled	The sum or the stock of words used by a language, a group, or an individual.
Comprehension [under 'Permute' index]	Comprehension; Language Tests	The act or fact of grasping the meaning, nature, or importance of; understanding. (American Heritage Dictionary, 4th ed) Includes understanding by a patient or research subject of information disclosed orally or in writing.
Pragmatics [under 'Stem' and 'Permute' indexes]	No MeSH term	No MeSH term definition
Grammar [under 'Stem' and 'Permute' indexes]	No MeSH term	No MeSH term definition
Morphology	No MeSH term	No MeSH term definition

Keyword	MeSH Terms found when searching keyword in PROSPERO [under 'Stem']	Definition of Keyword under MeSH
[under 'Stem' and 'Permute' indexes]		
Narration [under 'Permute' index]	Narration	The act, process, or an instance of narrating, i.e., telling a story. In the context of MEDICINE or ETHICS, narration includes relating the particular and the personal in the life story of an individual.
Phonology [under 'Permute' index]	Articulation Disorders	[Articulation Disorders definition] Disorders of the quality of speech characterized by the substitution, omission, distortion, and addition of phonemes.
Intervention	Clinical Trial; Intervention Study	[Definition of Clinical Trial] A work that reports on the results of a clinical study in which participants are assigned to receive one or more interventions so that researchers can evaluate the interventions on biomedical or health-related outcomes. The assignments are determined by the study protocol. Participants may receive diagnostic, therapeutic, or other types of interventions. While most clinical trials concern humans, this publication type may be used for clinical

Keyword	MeSH Terms found when searching keyword in PROSPERO [under 'Stem']	Definition of Keyword under MeSH
		veterinary articles meeting the requisites for humans.
Speech and Language Intervention [under 'Stem' and 'Permute' indexes]	No MeSH term	No MeSH term definition
Speech and Language Therapy [under 'Stem' and 'Permute' indexes]	No MeSH term	No MeSH term definition
Speech and Language Treatment [under 'Stem' and 'Permute' indexes]	No MeSH term	No MeSH term definition
Moderator [under 'Permute' index]	Effect Modifier, Epidemiologic or Moderator Variable or Moderator Variables or Variable, Moderator or Variables, Moderator	[Effect Modifier, Epidemiologic definition] Factors that modify the effect of the putative causal factor(s) under study.

Note. Stem finds terms that begin with that text. This is used when the permute index would bring back a very large list of irrelevant terms; Permute finds all terms that contain that text in any position. This is used when the term is so specific/ not a valid mesh term and will bring a similar number of terms/ not bring any terms back in the Stem index; Italicised terms are within one or more keywords.

Table 2.22. Stage 2: Search strings for findings duplicate reviews in PROSPERO

# String	Term	#Results
1	Child	5161
2	Child Behaviour	55
3	Child Behaviour Disorders	1
4	Child Development	229
5	Child Development disorders, Pervasive	0
6	Child Health	2968
7	Child Language	13
8	Child, Preschool	0
9	Children, Preschool	0
10	Preschool Child	32
11	Preschool Children	112
12	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11	5183
13	Developmental Language Disorder	10
14	Language Development Disorders	1
15	Developmental Language Disorders	2
16	#13 OR #14 OR #15	12
17	Oral Language	13
18	Expressive language	19
19	Receptive language	17
20	Vocabulary	1056
21	Language tests	3
22	Vocabulary Controlled	1

# String	Term	#Results
23	Comprehension	123
24	Pragmatics	13
25	Grammar	14
26	Morphology	230
27	Narration	32
28	Phonology	13
29	Articulation Disorders	5
30	#17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29	1452
31	Speech and Language Intervention	7
32	Speech and Language Therapy	52
33	Speech and Language Treatment	2
34	Intervention Study	487
35	Clinical Trial	3112
36	#31 OR #32 OR #33 OR #34 OR #35	3612
37	#12 AND #16 AND #30 AND #36	5
38	#12 AND #30 AND #36	29
39	Moderator	452
40	Effect Modifier, Epidemiologic	0
41	Moderator Variable	26
42	Moderator Variables	104
43	Variable, Moderator	0
44	Variables, Moderator	0
45	#39 OR #40 OR #41 OR #42 OR #43 OR #44	452

# String	Term	#Results
46	#37 AND #45	0
47	#38 AND #45	1

2.5.2. Appendix B. Planned synthesis, summary measures, data handling and combining data if meta-analysis and data pooling were viable

If participant data pooling or meta-analysis was viable, measures of treatment effect for initial intervention and ‘third variable’ analyses were planned to be presented differently for binary and continuous outcomes. Binary outcomes would be summarised using present or not present (0 or 1) values and will be estimated using odds ratio with a 95% confidence interval (Austin & Stuart, 2017). Continuous outcomes would be summarised using standardised mean differences in order to combine studies that measured the same outcome using different methods (Herbert, 2000). A fixed effects model would be used in the analysis if studies did not vary. However, if the data did vary due to potential study differences, a random effects model would be used (Borenstein et al., 2010). In the case of multiple outcomes and measures, this would be dealt with by meta-analysis methods that account for this such as multivariate meta-analysis (McKenzie et al., 2021). If more than one control group was utilised, the priority would be to analyse the main control (i.e., typically an alternative intervention), so that effect sizes were not ‘double counted’.

Sensitivity analysis were planned to be conducted for quantitative synthesis to determine if lower quality studies inflated outcomes or had different findings to higher quality studies (Tawfik et al., 2019). Funnel plots and trim and fill procedures would be used to assess publication bias, and forest plots were also planned to be used to examine heterogeneity (Tawfik et al., 2019). This would investigate the potential impact of differences in planned meta-regressions and sub-group analyses of factors. For assessing the publication bias of the included studies, funnel plots (Rothstein et al., 2005) were planned to be used to evaluate the relationship between effect size and published versus unpublished studies, and small versus large studies. If a relationship was identified, this was further examined for possible explanations. Each outcome was initially planned to be combined and calculated using STATA for meta-analysis and subgroup analysis.

2.5.3. Appendix C. Risk of bias evidence of decisions

Table 2.23. Cochrane risk of bias evidence for decisions: Selection random sequence generation and allocation concealment by study

Bias test	**1a. Selection random sequence generation**	**1b. Overall judgement**	**2a. Selection allocation concealment**	**2b. Overall judgement**
(1) Aguilar et al.	Receptive vocabulary scores were used to assign children to groups by ranking their scores and alternating assignment to each treatment condition. This was done to assure roughly equal vocabulary scores across groups. (p76)	High risk	N/A	N/A
(2) Bowyer-Crane et al.	Randomly allocated, but no details of method. (p423)	Unclear risk	Allocation was done in a way by a separate team member so that the researchers did not know. (p423 - see 4a.)	Low risk
(3) Dockrell et al.	No sufficient information to determine if this had occurred.	Unclear risk	No sufficient information to determine if this had occurred.	Unclear risk
(4) Goldstein et al.	Cluster randomized design with children nested in classrooms: classrooms were randomly assigned to the experimental and comparison conditions (p4): But no details on the randomisation process provided.	Unclear risk	No sufficient information to determine if this had occurred.	Unclear risk

Bias test	1a. Selection random sequence generation	1b. Overall judgement	2a. Selection allocation concealment	2b. Overall judgement
(5) Haley et al.	Randomly allocated to either the oral language intervention or the waiting control group. The randomization was conducted using an algorithm in Excel created by one of the contributing authors. (p73)	Low risk	The first author was initially blind to group membership, but this changed (see 4a). (p73) Implied other author was aware of groups.	High risk
(6) Justice et al.	Randomly assigned (p146) but no specifics provided.	Unclear risk	No sufficient information to determine if this had occurred.	Unclear risk
(7) Leonard et al.	No sufficient information to determine if this had occurred. "These assignments were made without regard to the children's ages or test scores" (p1366).	Unclear risk	No sufficient information to determine if this had occurred.	Unclear risk
(8) Motsch & Ulrich	Children were randomly assigned, but there is insufficient detail on how they did this. (p163)	Unclear risk	No sufficient information to determine if this had occurred.	Unclear risk
(9) Phillips et al.	Eligible children (n = 41) who were randomly assigned to the treatment condition were then assigned to 11 treatment subgroups comprising three to four children each (eight groups included four children). The non-random nature of the assignment was due to strategic decisions related to scheduling. (p1413)	High risk	N/A	N/A

Bias test	1a. Selection random sequence generation	1b. Overall judgement	2a. Selection allocation concealment	2b. Overall judgement
(10) Pollard-Durodola et al.	The researchers initially chose teachers from two school districts and randomly assigned them to one of two conditions (p165) The 148 students participating in the study were assigned to either treatment or comparison preschool classrooms on the basis of enrolment. (p166)	High risk	N/A	N/A
(11) Reeves et al.	Settings were randomly allocated (p57), but no indication of how this was done.	Unclear risk	No sufficient information to determine if this had occurred.	Unclear risk
(12) Smith-Lock et al. (a)	All of the 49 children were assigned to treatment conditions, as treatment was part of their regular classroom program. (p269) Each group contained four girls. (p269) Treatment conditions were allocated based on site (p271)	High risk	N/A	N/A

Bias test	1a. Selection random sequence generation	1b. Overall judgement	2a. Selection allocation concealment	2b. Overall judgement
(13) Smith-Lock et al. (b)	Children in their first year of full-time schooling were invited to participate in the study and assigned to treatment conditions, because treatment was part of their regular classroom program. (p314) Unclear how this was assigned	Unclear risk	Cluster randomization was used, with treatment randomly assigned by site (p314) One site randomly allocated the recasting procedure and one site randomly allocated the cueing procedure (p314) However, not said how this was done.	Unclear risk
(14) Van Kleeck et al.	The African American and the Caucasian groups of children were divided randomly into treatment and controls. To ensure random assignment, the children were assigned alternately to either the treatment or the control group as the children entered the study when their legal guardians returned the permission forms. (p88)	High risk	N/A	N/A

Bias test	1a. Selection random sequence generation	1b. Overall judgement	2a. Selection allocation concealment	2b. Overall judgement
(15) Wake et al. (2 papers)	Allocation done by an independent researcher by using a computer-generated random number sequence (p896) Randomization was stratified by previous trial (Let's Read or Let's Learn Language) and nature of language problem (receptive, expressive, or both receptive and expressive), and blocked within each stratum using randomly permuted block sizes in a non-systematic sequence (p897)	Low risk	Allocation concealed using sealed opaque envelopes (p896)	Low risk
(16) Washington et al.	Following parental consent, participants were consecutively assigned to C-AT or nC-AT (p318)	High risk	Allocation was not fully random and appears based on the knowledge of the researchers.	High risk
(17) Yoder et al.	To assign children to treatment groups, a computer program using a random number generator produced the random sequence. Even numbers were assigned to MLT and odd numbers were assigned to BTR. Participants were assigned to numbers in the order in which they were enrolled. (p6)	Low risk	The project director enrolled participants and was blind to treatment assignment at the time of participant enrolment. (p6)	Low risk

Table 2.24. Cochrane risk of bias evidence for performance bias and detection bias blinding

Bias test	3a. Performance bias blinding	3b. Overall judgement	4a. Detection bias blinding	4b. Overall judgement
(1) Aguilar et al.	Each clinician provided training with both activities, and different clinicians trained different children over the 3 days. This assignment was based on the availability of the clinician and the child. (p79)	High risk	The individuals testing the children did not participate in training, maintaining experimental blinding for both testers and training clinicians. (p78)	Low risk
(2) Bowyer-Crane et al.	To ensure that investigators were blind to group membership when testing, the research team was not involved in the allocation of children to the interventions. In the event of schools needing to contact the research team to talk about the programme, each school was allocated one member of the team as their contact. The other member of the team was assigned to carry out the assessments in that school. (p423-424)	Low risk	To ensure that investigators were blind to group membership when testing, the research team was not involved in the allocation of children to the interventions. In the event of schools needing to contact the research team to talk about the programme, each school was allocated one member of the team as their contact. The other member of the team was assigned to carry out the assessments in that school. (p423-424)	Low risk
(3) Dockrell et al.	Separate schools for each intervention, but it appears at least one school was aware of alternatives (due to being a non-intervention condition): Staff in the Non-intervention preschool received training in the Talking Time intervention after the study was finished, when post-intervention and data analysis were completed. (p505)	High risk	Assessment sessions were up to 30-min long. All assessors were trained psychologists, experienced with children and trained in the use of the psychometric tests. Assessors were blind to the intervention. (p502) -- but there was no information provided for how this was done.	Low risk

Bias test	3a. Performance bias blinding	3b. Overall judgement	4a. Detection bias blinding	4b. Overall judgement
(4) Goldstein et al.	Although classrooms were randomised, research staff assisted with any needs (p5), so likely aware of which classrooms were doing what interventions.	High risk	Research staff was responsible for the administration and scoring of child assessments (p5). Due to what is detailed in 3a, they were likely not blinded when assessing children.	High risk
(5) Haley et al.	Implied the waiting control group were aware of status as offered intervention after post-intervention: The waiting control group was offered intervention according to need after school entry; however, this was not monitored by the research team and was implemented at the discretion of each participating school based on their interpretation of their children's post-intervention performance and the overall programme effectiveness. (p73)	High risk	All testers were blind to group membership with the exception of the first author who conducted on-site tutorials where she observed an intervention session taking place, thereby gaining awareness of group membership before post-intervention testing occurred. (p73)	High risk
(6) Justice et al.	The researchers closely worked with intervention parents (e.g. helping train) (p147)	High risk	Researchers also measured outcomes. (p148)	High risk

Bias test	3a. Performance bias blinding	3b. Overall judgement	4a. Detection bias blinding	4b. Overall judgement
(7) Leonard et al.	Clinicians in the conditions were provided a list of sentence constructions to prevent them from using the other target as much as possible, and without them realising what the condition was. (p1370-71) They did produce less of other target, but there was no assessment to determine if they were aware of the condition.	Unclear risk	The judge who transcribed and scored the children's responses was not aware of the treatment condition to which a child was assigned and, consequently, did not know which morphemes, if any, constituted the targets. Because the judge helped administer the probes for some of the children, she often knew whether a given probe session represented the first or the second time the child had received the probes. However, she was also responsible for transcribing and scoring responses from audio recordings of probe sessions that she had not attended. In some of these instances she transcribed and scored responses from the posttreatment session before she transcribed and scored responses from the pretreatment session. Because these were audio recordings that lacked visual clues, it is likely that she was often unaware of the sequence in which these recordings were made. (p1368)	High risk

Bias test	3a. Performance bias blinding	3b. Overall judgement	4a. Detection bias blinding	4b. Overall judgement
(8) Motsch & Ulrich	The alternative intervention was no treatment, so the parents would be aware of children who were not receiving intervention. Unclear if researchers would consider this to be an effect on the results.	Unclear risk	To guarantee maximum objectivity of the results, the tests performed at T3 and T4 were 'single-blinded', meaning that the therapists performing the tests did not know which trial group the particular child belonged to (p164)	Low risk
(9) Phillips et al.	Control condition was no intervention. No detail provided to determine if all 5 school locations had intervention and control groups or not, and how they determined whether interventionists communicated or not. Either way, teachers would be aware of what condition they were in.	High risk	All post-intervention testing was conducted by assessors who were blind to children's treatment status and who had had no prior contact with them. Pre- and post-intervention data were anonymized and pooled prior to double scoring by blind scorers. (p1412)	Low risk
(10) Pollard-Durodola et al.	The researchers administered standardized and researcher-developed pretests and posttests to assess students' receptive and expressive vocabulary development. (p166), but no indication of blinding provided	Unclear risk	Trained graduate and undergraduate assistants individually administered all measures (p167), but no sufficient information to indicate whether they were blinded or not to intervention groups.	Unclear risk

Bias test	3a. Performance bias blinding	3b. Overall judgement	4a. Detection bias blinding	4b. Overall judgement
(11) Reeves et al.	The alternative intervention was no treatment, so the teachers would be aware of intervention or conrtrol assignment (p57). Unclear if researchers would consider this to be an effect on the results.	Unclear risk	The pre- and post-intervention assessments were carried out by speech and language therapists and speech and language therapy students under supervision who were blind to whether children had been in treated or control nurseries. Pre- and post-assessments for individual children were carried out by different assessors (p56)	Low risk

Bias test	3a. Performance bias blinding	3b. Overall judgement	4a. Detection bias blinding	4b. Overall judgement
(12) Smith-Lock et al. (a)	All of the 49 children were assigned to treatment conditions, as treatment was part of their regular classroom program. Of the 40 children tested for the study, 22 received treatment targeted at grammatical goals. Eighteen children received their usual treatment which focused on comprehension. (p269) Administration of the sites was the same (p271) Due to activities, teachers and clinician would be aware of what group children were in. (p271-272) By necessity, the speech pathologists and teachers were not blind to the intervention condition they were administering. The children involved in the study saw the treatment as a regular part of their classroom activities and had no contact with children in the other treatment condition. The children were very used to language instruction and regular testing and therefore could be considered blind to the entire process. (p273)	High risk	Three research assistants carried out the outcome testing. Each tester tested the same children in each testing phase to reduce the likelihood of test score changes being due to different testers. Two of the three testers were blind to the nature of the study. They were unaware that the children were participating in a treatment study, and by extension, were unaware of children's allocation to treatment conditions. The third tester, required due to last minute staffing issues, was not blind to the treatment conditions. (p273) A comparison of the gain scores of the children tested by the blind testers versus the children tested by the non-blind tester found no difference. (p277)	High risk

Bias test	3a. Performance bias blinding	3b. Overall judgement	4a. Detection bias blinding	4b. Overall judgement
(13) Smith-Lock et al. (b)	A double-blind superiority trial, but administration of the sites was the same (p314) Both testers and participants (children) were unaware of treatment conditions. The children involved in the study saw the treatment as a regular part of their classroom activities and had no contact with children in the other treatment condition. The children were accustomed to language instruction and regular testing and therefore can be considered unaware of the research process. By necessity, the teachers/ SLPs were aware of the intervention condition they were administering. (p317)	High risk	Two testers carried out the pre- and postintervention testing, and a third research assistant scored the tests. Each tester tested the same children in each testing phase, to reduce the likelihood of test-score changes being due to different testers. The testers and the scorer were unaware of the nature of the study. They were told they were studying grammatical development in the children, but they were unaware of the treatment component of the project. Poststudy interviews confirmed that the testers had remained unaware of the purpose of the testing. (p317)	Low risk
(14) Van Kleeck et al.	The alternative intervention was no treatment, so the research assistants would be aware of intervention or control assignment (p89). Unclear if researchers would consider this to be an effect on the results.	Unclear risk	Most of the testers were not blind to the children's group (treatment or control) status. This arrangement was necessitated by a combination of the number of personnel we had available to conduct this study and the constraints of their schedules and those of the Head Start children who participated. (p88)	High risk

Bias test	3a. Performance bias blinding	3b. Overall judgement	4a. Detection bias blinding	4b. Overall judgement
(15) Wake et al. (2 papers)	but once allocated, participants could not be blinded (p897); The control was no intervention. Unclear if researchers would consider this to be an effect on the results.	High risk	Outcome assessors were blind to group allocation (p897)	Low risk
(16) Washington et al.	Parents of children who were not receiving treatment (i.e., awaiting treatment) were asked to participate. This convenience sample of children served as control participants, no treatment (NT) (p318) To ensure treatment fidelity one clinician, the first author, a registered SLP, provided all intervention sessions (p321)	High risk	The language assessment batteries were completed by registered SLPs or graduate students supervised by registered SLPs. The IQ measure (i.e., theKBIT-2) was administered by the first author. Administration of the SPELT-P was completed pre-, post- and 3-months post-treatment by blinded assessors. (p319)	Low risk
(17) Yoder et al.	One clinician for each condition, so aware of what they were training and saw all children for that condition. (p7-8)	High risk	Same observers examined all data (p8-9)	High risk

Table 2.25. Cochrane risk of bias evidence for attrition bias and selective reporting

Bias test	5a. Attrition bias incomplete data	5b. Overall judgement	6a. Selective reporting	6b. Overall judgement
(1) Aguilar et al.	One child, originally in the high-variability condition, withdrew from the study after the training phase and did not complete the final experimental post-intervention tests. He was replaced with another boy who completed all study phases. Unclear why this was. (p75)	Unclear risk	Stated outcome and results for it.	Low risk
(2) Bowyer-Crane et al.	Attrition flow chart with numbers and reasons provided (p424). Certain events may have introduced bias (i.e. school withdrawing after allocation, 17 children being replaced following discussion with a teacher), no justifications or analyses provided to determine if this potentially affected results.	High risk	Not all pre (t1) post (t3) and follow-up (t4) values are reported as means and standard deviations. Full analyses not reported for all outcomes, and z-score bar chart difficult to decipher exact scores. States each outcome is for 67-72 participants, but does not specify ns for each outcome specifically.	High risk

Bias test	5a. Attrition bias incomplete data	5b. Overall judgement	6a. Selective reporting	6b. Overall judgement
(3) Dockrell et al.	It was not our original intention to consider only ELL children, and all children in each setting were given the pre-test measures and took part in the interventions. However, monolingual English-speaking children were unevenly distributed across the three settings, with only 8 of the 36 monolingual English speakers coming from the two settings where interventions were implemented. It was clear from the pre-test data that English monolingual children, despite performing at a low level with regard to oral language skills, performed significantly better than the ELL children on all language measures - analysed only the ELL children. We therefore decided to analyse data only from the 96 ELL present at post-intervention. This decision impacted most on the Non-intervention group, where ELL (17) and English monolingual (24) children were present in more equal proportions. It was not possible to conduct separate analyses of the	High risk	Stated outcomes and results for them.	Low risk

Bias test	5a. Attrition bias incomplete data	5b. Overall judgement	6a. Selective reporting	6b. Overall judgement
	performance of English monolingual children as by post-intervention there were no monolingual children in the Story Reading group, and only 4 in the Talking Time group Also stated other drop-out numbers, but no reasons why (p501).			

Bias test	5a. Attrition bias incomplete data	5b. Overall judgement	6a. Selective reporting	6b. Overall judgement
(4) Goldstein et al.	Attrition at the child level was accounted for with the addition of 15% more children, but attrition was expected to be unlikely at the classroom level. Did not explain this further (p4) No significant differences between the groups on demographic, developmental, or attrition variables (p4)	Unclear risk	Only reported differences between states for one outcome and not the other, and examined moderation for the experimental but not control group. (p8-9) English language learner status was not taken for one of the states, which could have had a clinically relevant impact on the results (Ohio) (p5)	High risk
(5) Haley et al.	Before randomization, one of the children originally selected wasexcluded due to the severity of her expressive speechand language difficulties (CELF Expressive Vocabularyscaled score of 0). This decision was made in consul-tation with the child's parent who contacted the firstauthor with concerns that the programme may not bethe right fit for her child. (p73) Have a flow chart of attritions with reasons clearly stated and unlikely to affect true outcome (5 lost due to moving schools). (p74)	Low risk	Stated outcomes and results for them.	Low risk

Bias test	5a. Attrition bias incomplete data	5b. Overall judgement	6a. Selective reporting	6b. Overall judgement
(6) Justice et al.	No sufficient information to determine if there was any attrition.	Unclear risk	Did not provide mean and standard deviation scores for composites, and zscore graph did not have exact numbers, so difficult to determine effect size. (p150-151)	High risk
(7) Leonard et al.	No sufficient information to determine if there was any attrition.	Unclear risk	Stated outcomes and results for them.	Low risk
(8) Motsch & Ulrich	Dropouts: At the time of T4 (12 months after completion of the intervention), two children from the CG and one child from the EG could not be tested further as they had moved away from the area or they had a long-term disease. (p164-165)	Low risk	Stated outcomes and results for them.	Low risk
(9) Phillips et al.	One child left his school after randomization but before pretesting, seven other children left subsequent to pretesting, and several were absent the week of midtesting. The analytic completer sample of children who received both pretesting and either midtesting, posttesting, or both included 77 children, for an attrition rate of just 6.1%. The eight children missing at posttesting were divided equally between the treatment and control group. (p1412)	Low risk	Stated outcomes and results for them.	Low risk

Bias test	5a. Attrition bias incomplete data	5b. Overall judgement	6a. Selective reporting	6b. Overall judgement
(10) Pollard-Durodola et al.	Of the 148 students (81 WORLD and 67 comparison) originally in the study, 23 (16%) dropped out before the study was completed. One teacher dropped out before the intervention began, another teacher opted not to participate during the intervention, and students typically dropped out because their families moved or because they withdrew from school during the school year. Of the 23 students lost to attrition, 12 were from the WORLD condition and 11 from the comparison condition. Chi-square analyses showed a nonsignificant difference in attrition rates (2 [1] = 0.07, p = .789) between groups. (p166-167), However, it was unclear if the demographics of the teacher or students differed between groups	Unclear risk	Stated outcomes and results for them.	Low risk

Bias test	5a. Attrition bias incomplete data	5b. Overall judgement	6a. Selective reporting	6b. Overall judgement
(11) Reeves et al.	Initially 18 nurseries volunteered to participate. Difficulties in scheduling the programme led to three nurseries dropping out. (p56) No indication of what these difficulties were and why drop-out occurred.	High risk	Stated outcomes and results for them.	Low risk
(12) Smith-Lock et al. (a)	All 49 children in their first year of full time schooling in one LDC were invited to participate in the study. Forty-five out of 49 students agreed to participate. Five of the children who agreed to participate were deemed ineligible due to diagnoses other than SLI. They seemed to take part in the intervention (as it was within selected classrooms), but were not tested at any point, so not removed post-hoc. (p269)	Low risk	Stated outcomes and results for them.	Low risk
(13) Smith-Lock et al. (b)	Description of drop-outs from original recruitment number provided with reasons (p314), but no indication of, or analysis to determine if this would affect outcomes	Unclear risk	Stated outcome and results for it.	Low risk

Bias test	5a. Attrition bias incomplete data	5b. Overall judgement	6a. Selective reporting	6b. Overall judgement
(14) Van Kleeck et al.	No sufficient information to determine if there was any attrition.	Unclear risk	Stated outcome and results for it.	Low risk
(15) Wake et al. (2 papers)	Diagram provided of attrition in groups with clear reasons. (p898) + levels of intervention received by participants (p900), but no indication of, or analysis to determine if this would affect outcomes	Unclear risk	Stated outcome and results for it.	Low risk
(16) Washington et al.	No sufficient information to determine if there was any attrition.	Unclear risk	Stated outcome and results for it.	Low risk
(17) Yoder et al.	IQ and occupational status factors differed between drop-out and analysed participants. (p6) = induced clinically relevant bias in the observed effect size	High risk	Outcome is clearly specified (e.g. p9-10)	Low risk

Table 2.26. Additional risk of bias for selective reporting of 'third variable' analysis

Bias test	7a. Selective reporting of ‘third variable’ analysis	7b. Overall judgement
(1) Aguilar et al.	Missing statistical information for analyses, but still interpretable. Unadjusted means only, so effect sizes could not be calculated. Only completed analysis for one of two outcomes (not justified). Same factors included in all outcomes.	High risk
(2) Bowyer-Crane et al.	Some uninterpretable findings due to missing statistical information. Unadjusted means only, so effect sizes could not be calculated. Analyses not completed for all outcomes (not justified). Same factors not included in all ‘third variable’ analyses for outcomes examined (not justified).	High risk
(3) Dockrell et al.	Some uninterpretable findings due to missing statistical information. Unadjusted means only, so effect sizes could not be calculated. Analysis not completed for all outcomes (not justified). Same factors included in all outcomes.	High risk
(4) Goldstein et al.	Some uninterpretable findings due to missing statistical information. Unadjusted means only, so effect sizes could not be calculated. Analysis completed for all outcomes. Same factors not included in all ‘third variable’ analyses for outcomes examined (not justified).	High risk

Bias test	7a. Selective reporting of 'third variable' analysis	7b. Overall judgement
(5) Haley et al.	Uninterpretable information - no reporting of covariate impacts. Adjusted marginalised means provided as part of a graph, but exact values hard to determine (and so effect sizes could not be calculated). Analysis completed for all outcomes. Same factors included in all outcomes.	High risk
(6) Justice et al.	Provided statistics for all analyses, but only provided unadjusted means, so effect sizes could not be calculated. Analysis completed for all outcomes. Same factors not included in all 'third variable' analyses for outcomes examined (not justified).	
(7) Leonard et al.	Missing statistical information for analyses, but still interpretable. Unadjusted means only, so effect sizes could not be calculated. Analysis completed for all outcomes. Same factors included in all 'third variable' analyses for all outcomes completed.	High risk
(8) Motsch & Ulrich	Missing statistical information for analyses, but still interpretable. Unadjusted means only, so effect sizes could not be calculated. Analysis not completed for all outcomes (not justified). Same factors not included in all 'third variable' outcomes completed (not justified).	High risk

Bias test	7a. Selective reporting of ‘third variable’ analysis	7b. Overall judgement
(9) Phillips et al.	Some uninterpretable findings due to missing statistical information. Adjusted means only – can calculate effect sizes, for ‘third variable’ but have no base comparison. Analysis completed for all outcomes. Same factors included in all ‘third variable’ outcomes completed.	High risk
(10) Pollard-Durodola et al.	Missing statistical information for analyses, but still interpretable. Unclear what type of means provided, but only either unadjusted or adjusted. Analysis completed for all outcomes. Same factors included in all ‘third variable’ outcomes completed.	High risk
(11) Reeves et al.	Missing statistical information for analyses, but still interpretable. Unadjusted means only, so effect sizes could not be calculated. Analysis not completed for all outcomes (not justified). Same factors included in all ‘third variable’ outcomes completed.	High risk
(12) Smith-Lock et al. (a)	Statistics provided for analyses. Unadjusted and adjusted means, so effect sizes could not be calculated. Analysis completed for all outcomes. Same factors included in all ‘third variable’ outcomes completed.	Low risk

Bias test	7a. Selective reporting of ‘third variable’ analysis	7b. Overall judgement
(13) Smith-Lock et al. (b)	Statistics provided for the analyses. Unadjusted means only, so effect sizes could not be calculated. Analysis completed for all outcomes. Same factors included in all ‘third variable’ outcomes completed.	High risk
(14) Van Kleeck et al.	Statistics not provided for analyses- uninterpretable. Unadjusted means only, so effect sizes could not be calculated. Analysis not completed for all outcomes (not justified). Same factors not included in all ‘third variable’ outcomes completed (not justified).	
(15) Wake et al. (2 papers)	Some uninterpretable findings due to missing statistical information. Unadjusted means and standard deviations only, but did have adjusted and unadjusted mean differences (these adjusted means reflected a combination of the factors). Mean differences could not help calculate effect sizes however. Analysis not completed for all outcomes (not justified). Same factors not included in all ‘third variable’ outcomes completed (not justified).	High risk
(16) Washington et al.	Statistics provided for any analyses completed.	High risk

Bias test	7a. Selective reporting of ‘third variable’ analysis	7b. Overall judgement
	Adjusted means only – can calculate effect sizes for factor but have no base comparison. Analysis not completed for all outcomes (not justified). Same factors included in all ‘third variable’ outcomes completed.	
(17) Yoder et al.	Missing statistical information for analyses, but still interpretable. Unadjusted means only, so effect sizes could not be calculated. Analysis completed for all outcomes. Same factors included in all ‘third variable’ outcomes completed.	High risk

Chapter 3. Identifying potential moderators of the relationship between early language and school readiness: Secondary data analysis of the Millennium Cohort Study

3.1. Background and research questions

3.1.1. The potential impact of child and social characteristics on the relationship between preschool language and school readiness

Chapter one highlighted that children's readiness to transition to formal schooling is associated with their long- and short- term educational and life outcomes (Davies et al., 2016; Jones et al., 2015; Law, 2015; Pan et al., 2019; Rahman et al., 2018; Reynolds et al., 2011; Ricciardi et al., 2021; Sadler et al., 2015). However, children differ in this ‘school-readiness’. School readiness is measured as a multi-component construct which typically is made up of components including academic, cognitive, behavioural, physical and socio-emotional skills (Camacho et al., 2019; Duncan et al., 2007; Kokkalia et al., 2019; Law, 2015; Pan et al., 2019). Furthermore, preschool language underpins and is associated with many of these components (Bretherton et al., 2014; Chow et al., 2018; Chow & Ekholm, 2019; Fuchs et al., 2018; Levickis et al., 2018; Lonigan & Milburn, 2017; Snijders et al., 2020; Trakulphadetkrai et al., 2020; Willinger et al., 2019; Yew & O’Kearney, 2013). Due to this, efforts to boost school readiness through early language is advocated for, and school readiness interventions targeting language have been widely used and successful in also improving other skills which comprise school readiness (EEF, 2019; Law et al., 2018; Lonigan et al., 2015; Nix et al., 2013; Noble et al., 2012; Welsh et al., 2020). But while intervention research demonstrates that preschool language interventions can benefit oral language and school readiness, their implementation and examination of effects tends to be motivated by set of implicit assumptions. Specifically, they assume 1) children benefit equally from language interventions (examined in chapter 2), and 2) children will benefit equally in school readiness outcomes from gains in language ability.

However, child and social factors also predict language intervention and school readiness outcomes (Betancourt et al., 2015; Duncan et al., 2015; Duncan et al., 2007; Flouri et al., 2020; Hosokawa & Katsura, 2018; Levickis et al., 2018;

Neuman et al., 2018; Paul, 2020; Vugteveen et al., 2021), so there may be numerous possible pathways or mechanisms through which oral language, school readiness, child and social factors associate with one-another. For example, previous research indicated that factors like SES may have an effect on school readiness independent of language, and may also affect the ability of children to capitalise on initial language advantages (Feinstein, 2003; Hammer et al., 2017; Prior et al., 2011). Why these findings may be of particular concern is that they could mean children get a ‘triple threat’ of disadvantage from developmental vulnerabilities and social disadvantage. Specifically, children with developmental vulnerabilities and social disadvantages could be receiving a cumulative disadvantage towards their language and school readiness development, gains from intervention, and gains in school readiness even if they benefit from interventions. If this does occur, then employing current interventions without considering how to tackle these levels of disadvantage will only compound difficulties that children have. Therefore, this research examines the potential moderating effect of child and social factors on the relationship between oral language and school readiness; which to my knowledge has not been explored previously. A set of hypotheses was made for each of the potential moderators considered.

3.1.2. Hypothesised moderating effects of child and social factors

In chapter 2, the findings from the systematic review were presented which indicated that some child and social factors not only directly affect preschool oral language growth, but may also produce differential intervention outcomes. Here I test the second hypothesis of this thesis that child and social factors could also influence the extent to which language ability benefits school readiness.

The inclusion of potential moderating factors was based on prior studies of the MCS and longitudinal research examining associations between child and social factors and school readiness outcomes (detailed below). Potential moderators included most of the factors which were also examined in the systematic review, with the addition of long-term health conditions, relative income poverty, home learning environment, childcare type, maternal age at child’s birth, and maternal mental health. The following presents hypotheses regarding the mechanisms through how factors might moderate the effect of language on school readiness with supporting evidence.

Hypothesised moderating effects of child factors

There have been a number of developmental vulnerabilities which have been identified to impact school readiness (age, gender assigned at birth, behavioural problems, speech difficulties, non-verbal IQ and long-term health conditions). The literature for each is presented, and then an overarching hypothesis is provided below.

Hobcraft and Kiernan (2010) found that for age, children in the MCS with summer birthdays (i.e., born earlier in the academic year) score lower in oral language and school readiness compared to their older peers. Furthermore, other longitudinal studies have found younger children demonstrated less socio-emotional development and performed worse in literacy and maths than older children in their class (Bassett et al., 2012; Murray & Harrison, 2011; Winsler et al., 2012; Ziv, 2013). These differences are likely because younger children have not had as much time to cognitively and socially develop compared to their older peers (Cantalini-Williams et al., 2016; Hobcraft & Kiernan, 2010).

In addition, a number of studies utilising MCS data found that being male predicted having poorer language, behaviour (measured with the SDQ), and school readiness outcomes at age 3 and 5 compared to being female (Camacho et al., 2019; Cullis & Hansen, 2008; George et al., 2007; Hobcraft & Kiernan, 2010; Mensah & Kiernan, 2010a). Additionally, gender impacts on school readiness outcomes were significant despite accounting for variables such as age, maternal age at birth, maternal education, SES, and if English was spoken at home (Quigley et al., 2012). This suggests that gender assigned at birth is a robust and direct influence upon school readiness. There are multiple suggestions as to why males and females could differ in school readiness skills. Research indicates that females' early brain maturation, cognitive (e.g., attention, executive function) and socio-emotional (e.g., emotion regulation, externalising behaviour) development is faster than males (Adani & Cepanec, 2019; Lovas, 2011; Maguire et al., 2016; Masnjak, 2017; Talbot, 2020; Unterrainer et al., 2013). This is shown in studies examining early developmental gender differences which show females express more advanced socio-emotional and communication skills (e.g., gesturing, eye contact, social referencing) and cognitive function (e.g., attention, processing speed) than males (Adani & Cepanec, 2019; Talbot, 2020). With this in mind, it appears that females are able to engage with, and access learning easier because they have more globally advanced development than males (Adani & Cepanec, 2019; Talbot, 2020).

Furthermore, studies examining MCS data and wider longitudinal research found behavioural problems (measured with the SDQ, or related to externalising and internalising problems) have been found to predict worse school readiness in both typically developing children and children with language disorder (Cullis & Hansen, 2008; Duncan et al., 2007; Hobcraft & Kiernan, 2010; Prior et al., 2011). Why children with behavioural difficulties perform poorer in school readiness could be because their behaviour is disruptive to their learning and engagement with academic tasks in the short-term, and missing out on the maximum possible benefit from learning makes them lag behind in the long-term (Patalay et al., 2016).

The systematic review in chapter 2 found that better speech increased intervention benefit and/or language growth. Research has also found that children with speech disorders/difficulties at preschool are more likely to have persistent difficulties throughout development in components of school readiness like behaviour, social communication and academic outcomes (Bishop et al., 2017; Davies et al., 2016; Roy & Chiat, 2014). Why children with different speech skills differ in school readiness outcomes could be because speech allows children to communicate their needs, and facilitates the development of their social skills (Simoni et al., 2019). Therefore, if children do not have good speech, then this may disrupt their socio-emotional development and ability to ask teachers for help, which in turn will disrupt their learning (Daniel & McLeod, 2017; Simoni et al., 2019).

Furthermore, the systematic review found that having a higher Non-verbal IQ (NVIQ) increased language growth. Longitudinal evidence for non-verbal cognitive skills was limited regarding school readiness, but research indicates that better working memory and executive function were found to predict better school readiness (Fitzpatrick & Pagani, 2012; Pellicano et al., 2017). Why children with different cognitive skills differ in their school readiness could be because working memory and executive function skills have been found to impact the development of multiple components of school readiness like academic skills and social-emotional conduct (Welsh et al., 2014). As such, if children are disadvantaged in cognitive processing, then they will lag behind in the development in components that make up school readiness, which in turn will negatively affect their school readiness outcomes.

Moreover, long-term health difficulties may also impact school readiness. Long-term health difficulties in children are defined by the Royal College of Paediatrics and Child Health (RCPCH, 2021) as incurable conditions which may be managed through treatment. They can be a wide range of medical issues, with some of the most common conditions for children being asthma, epilepsy, diabetes, cancer and learning disabilities. Children with long-term health difficulties are reported by the RCPCH to be much less likely to be enrolled in formal education (e.g. only 20% of children with an identified special educational need) and more likely to be separated from family and have disrupted school attendance due to their health condition (RCPCH, 2015). Nijhof et al. (2018) demonstrated that children with severe illness also face obstacles with play, which has important physical, emotional, cognitive, and social benefits. There has been little to no literature examining the effect of long-term limiting health difficulties for outcomes in the MCS, except for in Hobcraft and Kiernan (2010), which found that having a long-term health condition predicted poorer school readiness scores. When considering the evidence together, it could mean that having less access and ability to engage in educational settings and play for children with long-term health conditions could make them less able to be ready school.

From the evidence presented, it is hypothesised that developmental vulnerabilities (being younger, assigned male at birth, having lower scores in NVIQ, having behavioural problems, having long-term health conditions, or having speech difficulties) affects children’s gains in their school readiness over and above language, because they create developmental lags in multiple components of school readiness. The developmental differences in the other components of school readiness would then create barriers to how they access learning and engage with classroom activities and academic tasks.

Hypothesised moderating effects of social disadvantage

There have been a number of social disadvantages identified to impact school readiness. The social factors outlined here can be grouped into three main constructs, 1) income disadvantage (defined as disadvantage which relates to available resources of the family; poverty, home learning environment, and uptake of free school meals); 2) maternal factors (maternal education, maternal age at child's birth, maternal mental health); and 3) geographical/community disadvantage (defined as disadvantage relating to the child's location or community; area deprivation, childcare type). Each of these social disadvantage constructs link to school readiness, and literature and hypothesised interactions by each construct are highlighted below.

Income disadvantage

Examples of income disadvantage outlined in the current thesis are relative income poverty, home learning environment, and uptake of free school meals. Relative income poverty (referred to as poverty for the rest of the chapter) is a complicated phenomenon, but is generally defined as when an individual or family lacks income to meet a minimum standard of living in their country (Dickerson & Popli, 2016). This is important to differentiate from absolute poverty, which means individuals of families cannot afford food, shelter or education (Organisation for Economic Co-operation and Development, 2012). Relative income poverty was chosen as many MCS and longitudinal studies use a measure of relative poverty. The MCS dataset also does not have a measure specifically for absolute poverty. Poverty affects the ability to consistently afford important resources needed for language learning and educational development (Hansen & Kneale, 2013). Longitudinal research such as Isaacs (2012) found US children in poverty are less likely to be school ready compared to those in higher-earning families (48% compared to 75% respectively). In a study analysing the British Cohort Study, Feinstein (2003) found that children from families within the lowest quartile of income also demonstrated lower levels of educational attainment in adulthood compared to children from the middle or top income quartiles. This was the case even for children with initially higher educational scores. This indicates that children living in poverty are more negatively affected by educational development over time compared to more affluent peers. Other longitudinal research has indicated that children living in poverty and/or within lower income families scored lower in cognitive assessments (Beauregard et al., 2018), mathematics achievement (Johnson et al., 2022), and have a higher rate of behavioural problems (Hosokawa & Katsura, 2018) and emotion regulation issues (Chen & Miller, 2015) than children with more affluent parents. Furthermore, a study comparing the associations between fine motor skill development and family income at age 6 found worse development for children with lower family income compared to their more affluent peers (Aiman et al., 2016). In addition, studies analysing the MCS found poverty strongly predicts poorer performance in school readiness measures, and components of school readiness (oral language - expressive vocabulary, behaviour measured by the SDQ) at age 3 and 5, even when accounting for a multitude of other factors such as initial school readiness at age 3, age, maternal characteristics (e.g., depression, education, age at birth of child), and whether children lived in separated or single-parent households (Blanden & Machin, 2010; Camacho et al., 2019; Cullis & Hansen, 2008; Dex, 2008; Dickerson & Popli, 2016; George et al., 2007; Kiernan & Mensah, 2009; Mensah & Kiernan, 2010b, 2010a). Taken together, this means that relative poverty appears to be a unique and robust predictor for school readiness outcomes regardless of the presence of other characteristics.

Furthermore, the Home Learning Environment (HLE) has been acknowledged as an important contributor to school readiness for over 30 years (Elardo & Bradley, 1981). Melhuish et al. (2008) found that higher scores of HLE positively predicted early language and educational outcomes. Niklas and Schneider (2017) provided support for Melhuish et al.’s (2008) study, finding that in a large-scale longitudinal study of 900 German children, a better HLE not only predicted academic higher competencies at the beginning of school, but also higher academic achievement at nearly 10 years old. Other large-scale studies have also found evidence that a higher-quality HLE predicts better outcomes in abilities relating to school readiness (Kluczniok et al., 2013; Rodriguez & Tamis-LeMonda, 2011; Son & Morrison, 2010). Furthermore, Cullis and Hansen (2008), and Hobcraft and Kiernan (2010) found that when parents spent less time reading to children or practicing alphabet and counting with them (all considered aspects of HLE) this predicted worse school readiness outcomes.

Finally, literature identifies children receiving free school meals as a proxy for social disadvantage, and find children who are socially disadvantaged are at risk of being less school ready (Illøkken et al., 2021; Sørensen et al., 2016; Winsler et al., 2008).

From the evidence presented, why income disadvantage affects children's gains in their school readiness over and above language could be because children are more likely to have barriers to educational resources and enriching learning experiences that improve cognitive, socio-emotional and academic development (Duncan et al., 2014; Hobcraft & Kiernan, 2010; Illøkken et al., 2021; Mollborn et al., 2014). Therefore, if children had less access to these resources, this in turn could reduce their knowledge of the world within which to contextualise their language learning and/or their familiarity, and therefore engagement with more formal learning activities.

Maternal factors

The maternal factors highlighted here are maternal education, maternal age at birth, and maternal mental health. MCS and other longitudinal research has consistently found that children with mothers with more qualifications are likely to have better preschool language, cognitive skills, behaviour (measured by the SDQ) and school readiness outcomes than mothers with fewer qualifications (Camacho et al., 2019; Cullis & Hansen, 2008; George et al., 2007; Harding et al., 2015; Hobcraft & Kiernan, 2010; Hosokawa & Katsura, 2018; King et al., 2017; Magnuson et al., 2009; Montroy et al., 2019; Reid & Strobino, 2019).

Furthermore, Tearne (2015) found in their review that children were less at risk for worse behavioural and academic outcomes if they had older mothers compared to mothers in their teens or twenties. There were also some studies analysing the MCS that found maternal age may be impactful on language and school readiness outcomes. Morinis et al. (2013) found that having an older mother also predicted better expressive vocabulary and cognitive scores. This was regardless of a range of maternal (psychological distress, self-esteem, attachment), birth (hospital days post-partum; children in the household; whether the child was the first-born) and sociodemographic factors (e.g., parental income and maternal education level), parenting, childcare and HLE. Additionally, Sutcliffe et al. (2012) found evidence that expressive vocabulary outcomes at age 3 and 5 were worse when their mothers were teen-aged. Studies analysing the MCS and examining maternal mental health, children's lower attainment on expressive vocabulary and school readiness was associated with higher rates of parental psychological distress (Cullis & Hansen, 2008; Hobcraft & Kiernan, 2010; Mensah & Kiernan, 2010b). Furthermore, Papachristou & Flouri (2020) found when using latent growth curve modelling on MCS data that both maternal age and maternal mental health contributed to the variation in children's early school readiness, and in the trajectory of behaviour (measured using the SDQ) and expressive vocabulary development from preschool to adolescence. Furthermore, Hobcraft & Kiernan (2010) found children born from very young mothers had a higher risk for poorer school readiness scores at age 5.

Why maternal factors affect children's gains in their school readiness over and above language may be because they predict the availability of educational resources available to 'invest' in their children's academic and socio-emotional development (Clifford et al., 2021; Dickerson & Popli, 2016; Goisis et al., 2017; Harding et al., 2015; Hosokawa & Katsura, 2018; Jackson et al., 2017; McDonald et al., 2016). This in turn could mean children are likely to develop less in the various components underlying school readiness, which results in them being less able to access learning and engage with formal learning activities.

Geographical disadvantages

Examples of geographical disadvantages outlined in the current thesis are area deprivation and childcare type. Area deprivation is a measure which ranks neighbourhoods by their lack of financial and social resources, crime, and safety of residents (Flouri et al., 2020). When area deprivation was used as a predictor in analyses of the MCS, Hobcraft and Kiernan (2010) found that children living in deprived areas were less likely to be school ready. While only a component of school readiness, Flouri and colleagues found increases in behavioural problems (measured by the SDQ) were both predicted by a more general measure of neighbourhood deprivation (including area deprivation, Flouri et al., 2012) and specifically by area deprivation (Flouri, 2020).

Why area deprivation affects children’s gains in their school readiness over and above language may be because the lack of communal material and social resources directly impact poor social control in the community, influences parental depression, and has a lack of social opportunities (Flouri, 2012; 2020). These issues then may make children more vulnerable to developing behavioural problems (Flouri, 2020). When theorising behaviour previously, having behavioural problems is disruptive to children’s learning and engagement with academic tasks in the short-term, which could lead them to missing out on the maximum possible benefit from learning and lagging behind in the long-term (Patalay et al., 2016).

In addition to area deprivation, childcare type was also a notable factor of interest. Childcare type is defined as a setting or service where children are cared for (Roberts et al., 2010). Different types include formal settings like nursery or schools, and informal types include family and friends (Del Boca et al., 2018; Hobcraft & Kiernan, 2010). Large-scale and longitudinal studies and reviews have found evidence that childcare quality is positively associated with better and long-term cognitive and educational outcomes; but quality varies across childcare settings (Bernal & Keane, 2011; Brilli et al., 2013; Burchinal et al., 2015; Del Boca et al., 2018; Gregg et al., 2006; Gregoriadis et al., 2016; Hiilamo et al., 2018; Li et al., 2013; E. Melhuish & Gardiner, 2018; National Institute of Child Health and Human Development, 2006; Paull et al., 2002). One study analysing the MCS found that not having attended any pre-school care predicted worse outcomes in school readiness scores (Hobcraft & Kiernan, 2010). Furthermore, Hopkin et al. (2009) found attending a formal preschool setting significantly and positively affected school readiness outcomes, while more informal childcare types had the opposite effect. As such, children receiving formal childcare could have educational advantages that allow them to contextualise their language learning effectively, and learn and engage with formal schooling more effectively over children who receive informal childcare.

3.1.3. Choosing secondary data analysis of the Millennium Cohort Study

To assess the hypothesis that all children will benefit equally in school readiness outcomes form gains in language ability, it was important to consider carefully the data needed. Data would have to place language at a separate time before school readiness, to ensure that it could be established that language predicted school readiness. In addition, data for the candidate moderators would have to be measured at the same time as language or before. This is so a clear interaction could be assumed with language (Montoya, 2019). Measures would also have to be high quality and represent a spectrum of backgrounds/ability. This is to ensure hypotheses could be explored, the research question could be adequately addressed, and conclusions would be valid and robust (Price et al., 2019). Furthermore, the sample needed to be large enough and to be able to identify potentially small effects and adjust for potential confounders (Kahlert et al., 2017; Matz et al., 2017). For these reasons, secondary analysis of data from the Millennium Cohort Study (MCS) was therefore chosen as it meets all of the criteria previously outlined.

The MCS is currently one of the most comprehensive and largest longitudinal studies of development in the world. It is a nationally representative birth cohort study currently maintained and conducted by the Centre for Longitudinal Studies (CLS). Data collection for the MCS began in 2000 and obtained data in ‘waves’ every 2-4 years using a variety of cognitive, social, psychological and economic measures (Joshi & Fitzsimons, 2016). These were collected via health assessments, cognitive tests, observations, questionnaires and interviews from cohort children, family members and teachers (Joshi & Fitzsimons, 2016). The wealth of data collected has been used widely in research to examine how individual and social circumstances influence outcomes in education, mental health, behaviour and cognitive abilities from early in development, (e.g. Camacho et al., 2019; Hansen et al., 2010). For the purposes of this study therefore the MCS provided a measurement of language at a time period before school readiness was measured; high-quality, rich data for factors of interest; and a population representative sample.

3.1.4. Research aim

To examine whether children benefit equally in their school readiness outcomes from gains in language ability, or if these benefits are moderated by additional child and social factors.

To do this, the study had two phases:

1) To identify key moderators to examine in the MCS, while considering issues of both theory and methodology;
2) To examine the effect of the identified moderators on the effect of language on school readiness

3.2. Method

3.2.1. Ethical approval and dataset access

Ethical approval for this phase was granted from Newcastle University's Humanities and Social Sciences Ethics Committee in December 2017. Additional conditions of use (confidentiality) by the UK Data Service were accepted to use the Millennium Cohort Study datasets in May 2019.

3.2.2. Study design

Waves chosen from the Millennium Cohort Study

The MCS has eight waves and has to date collected data from 9 months to 22 years. The waves chosen were based on when the key outcome of interest (school readiness at the beginning of school) was collected, and any relevant waves before that (i.e., had language measures and relevant moderators and covariates). Age 5 (wave 3) was chosen as this is when a high-quality school readiness measure, the Foundation Stage Profile, was employed. The measure was nationally used and moderated by the government, as well as measuring a range of school readiness abilities (further details outlined in the below section and in appendix E). Age 3 (wave 2) was chosen for predictors, moderators and covariates as it was the first wave to collect oral language data, and had several relevant variables to choose from. The first wave (9 months) did not have oral language data, and waves 4 onwards (age 7 +) would have been too late to obtain school readiness data.

Predictor and outcome variables: Measures of language and school readiness

Predictor variable (language): British Ability Scales 2 Naming Vocabulary

Expressive vocabulary, measured by the British Ability Scales 2 Naming Vocabulary subtest (BAS-2 Naming Vocabulary, Elliott, 1996) was selected as the predictor variable. The BAS-2 Naming Vocabulary has 36 items (including 2 teaching items) where children are shown pictures of objects and asked to name them. This was the only measure for oral language in wave 2, meaning no other expressive skills or receptive oral language measures were used. However, the BAS-2 Naming Vocabulary assessment is considered a high quality and valid measure for assessing expressive vocabulary (Connelley, 2013). Further information about this measure is reported in appendix E.

Outcome variable (school readiness): Foundation Stage Profile

As outlined in chapter 1, school readiness is a multi-component construct. In the assessment used in the MCS, the Foundation Stage Profile (FSP) explicitly examined six areas of learning relating to the English curriculum and covering children's physical, intellectual, emotional, creative and social development (Department for Education and Skills, 2003; Hansen & Jones, 2008). The FSP is completed for children who will be 5 years old on, or before, 31 August of that academic year (Department for Education and Skills, 2003). This assessment was only completed in England as opposed to the other UK countries (Scotland, Northern Ireland and Wales). There are 6 main assessment scales (with some being split into further subscales) which represent six overall areas of learning (subdivided into 13 areas of learning):

1) Personal, social and emotional development: a) dispositions and attitudes, b) social development and c) emotional development);
2) Communication, language and literacy: a) language for communication and thinking, b) linking sounds and letters, c) reading and d) writing
3) Mathematical development: a) numbers as labels for counting, b) calculating and shape, and c) space and measures;
4) Knowledge and understanding of the world;
5) Physical development; and

6) Creative development.

A child’s readiness in these areas of learning is assessed by whether they demonstrate specific early learning goals (ELGs). These are behaviours relevant to the areas of learning. Some examples of ELGS are “Shows an interest in classroom activities through observation or participation.” (p.6, personal, social and emotional development; Department for Education and Skills, 2003), and “Uses language to imagine and recreate roles and experiences.” (p.20, communication, language and literacy; Department for Education and Skills, 2003). If a child demonstrates an ELG, then they are scored a point on the FSP. Each of the 13 subscales have 9 ELGs each (so 27 ELGs for personal, social and emotional development; 36 ELGs for communication, language and literacy; 27 points for mathematical development; and 9 ELGs each for knowledge and understanding of the world, physical development and creative development). Therefore, each subscale has a maximum of 9 points, and scales (the overall areas of learning) have a maximum score between 9 and 36 points. The first 3 ELGs reflect children progressing towards achieving an area of learning (these are expected to be obtained by most children). ELGs 4-8 expected achievement in an area of school readiness. The ninth ELG reflects a child who has achieved all of the previous eight points, and indicates ability beyond the level of the early learning goals (Department for Education and Skills, 2003). Therefore, higher scores would indicate more mastery of a specific area. Scoring a point should reflect that the child is consistently performing an ELG by the end of the reception year. The individual requirements a child needs to demonstrate to be credited with each ELG can be found in the FSP handbook (Department for Education and Skills, 2003).

In addition to using total score, a derived variable for school readiness was created to reflect the government standard – the Good Level of Development (GLD) score. GLD is achieved if a child scores an average of 6 or more across the 7 ‘Personal, social and emotional development’ and ‘Communication, language and literacy’ subscales; in addition to scoring 78+ points overall. There was also no government documentation for determining GLD at the wave year the children were assessed, and so a more modern version of documentation was used (Department for Education, 2010). Before 2010 (the MCS children were assessed in 2004), GLD only required a total score of 78 or more on the FSP. This threshold likely changed to also requiring ‘average’ scores for personal, social and emotional development and communication, language and literacy scales due to the emphasis on communication, language and socio-emotional development goals of the government (Early Years Matters, 2008). So, while different to the GLD requirements at the time children were assessed, the choice was to utilise the more conservative threshold to reflect current practice and support people to apply to current populations.

Sampling frame of the MCS

Detailed information about the sampling frame, recruitment, productivity and attrition rates are provided in the MCS user guide (Centre for Longitudinal Studies, 2020). Briefly, the MCS obtained ethical approval for each study wave from NHS Research Ethics Committees (Shepherd & Gilbert, 2019). The sampling frame for the MCS included families which were eligible to receive Child Benefit (register provided by the Department of Social Security, and Department for Work and Pensions (DWP)). They also had to live in one of the four UK countries (England, Scotland, Wales and Northern Ireland). Births sampled for the cohort were across a 16-month period rather than a week or month like in most prior birth cohorts. This was to allow for easier, less intensive data collection; in addition to having the possibility to examine season-of birth effects. The full MCS sample were stratified into three strata; 1) ‘ethnic minority’ stratum (at least 30% proportion of a ward was populated by people with an ethnic minority status according to the 1991 census – England only), 2) ‘disadvantaged’ (in the poorest 25% of wards according to the Child Poverty Index for England and Wales but excluding those in the ‘ethnic minority’ stratum), 3) ‘advantaged’ (living in wards which are not defined by 1 or 2). The ‘ethnic minority’ stratum was only included for England as the other 3 UK countries did not have a very high ethnic minority population. The sample was clustered by characteristics of electoral wards for data collection efficiency and to allow for the examination of area effects.

Dataset and variable set-up

Relevant datasets were acquired from the UK data service after consulting Centre for Longitudinal Studies (CLS) documentation. Once obtained, merging, data cleaning and re-coding of variables (where appropriate for analyses) was completed in STATA software version 16. See appendix D for a detailed overview, and STATA syntax.

3.2.3. Phase 1: Variable selection process

Theoretical and methodological considerations

The research aim was addressed in two phases. The first was to identify key moderators to examine in the MCS, while considering the robustness of both theory and measures. As an exploratory analysis forming part of a doctoral thesis, it was important to keep the scope of the analyses manageable. Therefore, one each of the best quality child and social characteristics were selected to explore as potential moderators. Decisions to include factors as moderators/covariates are summarised in Table 3.1, and further information relating to the selection process for each variable is reported in appendix E.

Variables were selected as potential moderators based on their availability, and theoretical fit (i.e., association with school readiness, how it could moderate the relationship between preschool language and school readiness, links with other factors which may affect its viability as a moderator). It was important to establish the child and social factors potential importance to the predictor and outcome as based on previous literature reviewed above (in the background section), and that it was theoretically plausible that they could moderate the relationship. Next, methodological (measure quality) considerations were made to decide whether to include as moderators, covariates, or to exclude. As outlined previously, language development and school readiness have been shown to be predicted by the child and social factors considered for moderation analysis. As such, they could potentially distort findings and the true effect of selected moderators. Therefore, variables not chosen as moderators which were also methodologically robust were chosen as covariates. Some factors were completely excluded if the quality of the measure/data was too poor (i.e., speech difficulties, childcare type, long-term health conditions), were not available as measures in the MCS (i.e., Non-verbal IQ) or technically could not be put in the moderation model (i.e., area deprivation).

The predictor, covariates and moderators were selected from wave 2, while the outcome was selected at wave 3. Sources of information to examine variables were a) MCS documentation, data dictionaries (provided with the dataset), measure/ assessment documentation, and literature; b) The systematic review data obtained in phase 1; and c) their initial descriptive statistics.

Results

The resulting choices for moderators were gender assigned at birth (male/female) and relative income poverty (OECD above/below 60% poverty median). The covariates were initial school readiness (Bracken School Readiness Assessment-Revised), age (in months), behaviour (SDQ), home learning environment (HLE Index), maternal education (NVQ level), maternal age at birth (in years and months) and maternal mental health (Kessler-6).

Table 3.1. Factor roles selected for analysis

Variable	Candidate roles	Final role selection reasons	Final roles
Child factors			
Initial language (expressive vocabulary – British Ability Scales 2 Naming Vocabulary)	Predictor	Only available language measure (theoretically important to the research question). Expressive vocabulary is a commonly utilised measure of preschool language. Has excellent validity, reliability, representativeness, and data quality; has good generalisability (it was decided by assessors/parents that some children were unable to take the assessment due to not speaking English, or having a disability that made it inappropriate to take test).	Predictor
Gender assigned at birth (Male/Female)	Moderator or covariate	Related to language development and robust and direct influence on school readiness. Has not been explored before as a moderator in good quality research (as seen in the systematic review especially), so would be novel to examine. Has excellent validity, reliability, representativeness, and data quality.	Moderator
School readiness (Foundation Stage Profile)	Outcome	Covers a broad range of school readiness skills that can be separated and formulated into a government mandated threshold (i.e., GLD). Theoretically important to the research question. Has excellent validity, reliability, representativeness, data quality; good generalisability (English speakers only).	Outcome

Variable	Candidate roles	Final role selection reasons	Final roles
Age (in months)	Moderator or covariate	Evidence of being related to school readiness, and showed potential as an influential factor for language outcomes in the systematic review. But is difficult to separate conceptually from language and school readiness (e.g., it is not the language skill, but actually the age that may influence school readiness). Has excellent validity, reliability, representativeness and data quality.	Covariate
Non-verbal IQ	Moderator or covariate	Some evidence of being related to school readiness. Also showed mixed potential as an influential factor for language outcomes in the systematic review. No variables for this were available in MCS dataset.	Excluded
Comorbidity – Speech (Speech and language concerns)	Moderator or covariate	Demonstrated that it could be an important moderator for language intervention response in the systematic review, and related to persistent difficulties in school readiness skills. Only variable which indicated speech difficulties was of poor quality (vaguely labelled categories without definitions, mixed in with language concerns).	Excluded
Comorbidity – Behaviour (SDQ)	Moderator or covariate	Related to school readiness (for children with and without language disorder). Has excellent validity, representativeness, and reliability; issue with generalisability (over-representation of behavioural difficulties compared to UK estimates, likely due to being a screening rather than diagnostic measure), and data quality because it has a moderate amount of missing data (n=454).	Covariate

Variable	Candidate roles	Final role selection reasons	Final roles
Comorbidity – Long-term health condition status (yes/no and if it limited their everyday activities)	Moderator or covariate	Very little research conducted, but suggestion that having a long-term condition impacts school readiness. Also chosen based on the potential that medical issues could impact school attendance and play (which contributes to developing school readiness skills). Has issues with validity, reliability, representativeness, generalisability and data quality. Was unclear what it measured, what subgroups were present, and unclear what it would mean for policy and practice it if it was significant.	Excluded
Initial school readiness (Bracken School Readiness Assessment - Revised)	Covariate	It has been used as a precursor to the FSP assessment in longitudinal MCS studies, and strongly predicts its outcomes, or is used to help account for it. Has excellent validity, representativeness and reliability; has good generalisability (assessed English speakers only); issue with data quality as it has a moderate amount of missing data (n=494).	Covariate
Social factors			
Relative income poverty – (Organisation for Economic Co-operation and Development relative poverty 60% threshold)	Moderator or covariate	Related to language development and robust, direct and persistent influence on school readiness. Has not been explored before as a moderator in good quality research (not examined at all in the systematic review), so would be novel to examine. Has excellent validity, reliability, and data quality; has fairly good representativeness (number of children living above the poverty threshold in sample slightly higher than expected according to the measure estimates).	Moderator

Variable	Candidate roles	Final role selection reasons	Final roles
Socio-economic status – maternal education (NVQ levels)	Moderator or covariate	Related to language development and school readiness. But may be more of an indirect variable, and so may mediate some of the other variables (e.g., availability of resources, parental mental health, etc.). As such, it may be best to account for these so effects of other variables are not due to this. Has excellent validity, reliability, representativeness, and data quality; has good generalisability (children whose mothers had overseas qualifications were removed, but common practice in MCS data analysis and due to their vague description).	Covariate
Socio-economic status – area deprivation (Index of Multiple Deprivation: Living Environment)	Moderator or covariate	Related to language development and school readiness (findings are both separate and as part of the Index of Multiple Deprivation measure), and showed potential as an influential factor for language outcomes in the systematic review. This was not chosen over poverty as a moderator because it was not as well established in the literature for school readiness outcomes. Has excellent validity, reliability, representativeness and data quality. However, no findings could be obtained for the analyses when it was MCS weighted. This may be due to the weights having an emphasis on wards, and so area data may already be accounted for which means including this variable in the adjusted analyses was inappropriate.	Excluded

Variable	Candidate roles	Final role selection reasons	Final roles
Socio-economic status – Free/reduced school meal uptake	Moderator or covariate	Some evidence of being related to school readiness, and showed its potential as an influential factor in the systematic review. No variables for this were available in MCS dataset.	Excluded
Childcare type	Moderator or covariate	Related to cognitive and educational outcomes (including language) and school readiness. Could not create a derived variable which would be valid or reliable as the quality of data was poor (e.g., unclear or vaguely described categories, confusing variables used for data in dataset).	Excluded
Home learning environment (HLE index)	Moderator or covariate	Related to language development and school readiness. However, it may be difficult to separate from socio-economic factors, as better ‘investment’ in better learning environments and activities have been shown to be dependent on resources, and so it may be hard to determine if effects found are based on it or other variables. Has excellent validity, reliability, and data quality. Also has good representativeness (unclear what specific populations were used to ‘norm’ the measure, but was developed from assessing children in multiple preschool centres).	Covariate
Maternal age at birth (in years)	Moderator or covariate	Related to language development and school readiness. However, seems to situated in a complex system of other variables (e.g., poverty), and so it may be hard to determine if effects found are based on it or other variables. Has excellent validity, reliability, representativeness and data quality.	Covariate

Variable	Candidate roles	Final role selection reasons	Final roles
Maternal mental health (Kessler-6)	Moderator or covariate	Related to language development and school readiness. However, it may be mediated by, or closely linked to maternal education, and so it may be hard to determine if effects found are based on it or maternal education. Has excellent validity,and reliability; has good representativeness (data was skewed to lower scores, therefore may be underrepresenting mental health difficulties, but this could depend on the prevalence data examined); issue with data quality is that it has a moderate amount of missing data (n=745).	Covariate

Note. Generalisability is the degree to which the measure was able to be used across the whole sample. Data quality relates to amount of missing data and/or how useable/clear data was to use and analyse.

3.2.4. Phase 2: Moderation analysis

Research questions

The second phase of the study was to examine the effect of key potential moderators (gender assigned at birth and poverty) on the effect of language on school readiness.

The following research questions were addressed:

1) To what extent is expressive vocabulary at age 3 associated with a child's readiness for school at age 5?
2) Does gender assigned at birth alone and after adjustment for potential confounders moderate this effect? If so to what extent?
3) Does relative poverty alone and after adjustment for potential confounders moderate this effect? If so to what extent?

Measures

It was decided to analyse the outcome in three ways:

1) Foundation Stage Profile (FSP) total: total score across the 6 scales (out of 117), a continuous outcome which reflects a spectrum of school readiness ability.
2) 'Good Level of Development' (GLD): a binary variable (achieved/not achieved) derived using government identified threshold (average score of 78+ overall, and 6 or more across each of the 7 'Personal, social and emotional development' and 'Communication, language and literacy' scales).
3) Each FSP scale: An issue with both the total and GLD scores is that they are made up of subscales consisting of more points and categories than others (i.e., Personal, social and emotional development and Communication, language and literacy and Mathematical Development scales). One scale ('Communication, language and literacy') strongly relates to oral language as it includes language and communication skills. This could mean there is a possibility that significant or positive effects would be due language predicting a measure with an emphasis on

language. As such, it was important to have analyses for each sub-scale and compare them to the findings for the FSP total score and the GLD binary variable.

The language predictor was expressive vocabulary (BAS-2 Naming Vocabulary T-score). Covariates in models 4 and 5 were initial school readiness (Bracken School Readiness Assessment-Revised standard score), age (in months), behaviour (SDQ total difficulties), home learning environment (HLE Index total score), maternal education (no qualifications to NVQ level 5), maternal age at birth (in years and months) and maternal mental health (Kessler-6 total score). The variable not used as a moderator was also used as a covariate in the opposite model (i.e., gender assigned at birth (male/female) used as a covariate in the poverty (above/below poverty threshold) moderator model, and vice versa).

Initial analyses

Descriptive statistics by child and social factors are reported. In addition, mean/proportions, SD/SE, ranges and CIs and group comparisons (via t-tests and chi-square analyses) were calculated for the predictor, and outcome overall and for each level of the moderator. This was to gauge if subgroups had initial differences, and to help with the interpretation of later moderation analyses.

Main analyses

Regression analyses were chosen as they are suitable for including multiple variables, and can demonstrate each variable's individual association (coefficients), their collective variance, and include interaction analyses (Brook & Arnold, 2018). As some of the covariates are linear, regression allows for a more interpretable understanding of change per the unit of a variable unlike other interaction analyses types like ANCOVAs (Brook & Arnold, 2018). Depending on the outcome, linear (FSP total, FSP scales are continuous) and logistic (FSP GLD is binary) regressions were conducted (e.g. Su et al., 2012). Assumptions were tested for both linear and logistic regressions (Kasza & Wolfe, 2014; Schreiber-Gregory & Bader, 2018). For the linear regressions, this was 1) linearity between independent and outcome variables (examined via scatterplot), 2) normality of residuals (examined via kernel density, P-P and Q-Q plots), 3) no multicollinearity (via VIF and Pearson's correlations), 4) no homoscedasticity (examined via scatter plot). The assumptions for logistic regression were 1) outcome is binary, 2) observations need to be independent of each other, 3) no multicollinearity (via Spearman's correlations), 4) linearity between independent variables and log odds (via Box-Tidwell test), and 5) large sample size. For each outcome type, five regression models were analysed both unadjusted, and adjusted for MCS weighting:

1) Expressive vocabulary (predictor) and school readiness (outcome);

2) Expressive vocabulary (predictor) x gender assigned at birth (moderator) and school readiness (outcome);

3) Expressive vocabulary (predictor) x poverty (moderator) and school readiness (outcome);

4) Expressive vocabulary (predictor) x gender assigned at birth (moderator); poverty, initial school readiness, age, behaviour, home learning environment, maternal education, maternal age at birth, and maternal mental health (covariates); and school readiness (outcome); and

5) Expressive vocabulary (predictor) x poverty (moderator); gender assigned at birth, initial school readiness, age, behaviour, home learning environment, maternal education, maternal age at birth, and maternal mental health (covariates); and school readiness (outcome)

Models were completed and compared separately to determine how much the additional variables add to the base model, and if this affected the fit of the model. As all analyses were complete case analyses, the number of children in the sample also changed depending on model. Coefficients, R^2 and p values were provided alongside scatter plots. Standardised beta values are not provided, as the moderators are not standardised measures and could not be interpreted effectively via standard deviation changes (Hayes, 2017). Unstandardised betas are on the other hand are advised in most cases for clearer interpretation (Hayes, 2017). Findings should therefore be interpreted as relating to a one-unit (T score) change in expressive vocabulary and its effect on *n* unit changes in FSP (raw score). For the moderator gender assigned at birth, the reference values are being male and for poverty below the poverty line. Hence beta values can be interpreted as the raw score (total/scales) or likelihood of achievement (GLD) females and children living above the poverty line gained compared to males and children living below the poverty line respectively. Figures 3.1 and 3.2 illustrate the final moderator models with covariates (i.e., moderation models 4 and 5).

Figure 3.1. Child characteristic (gender assigned at birth) moderator model (model 4)

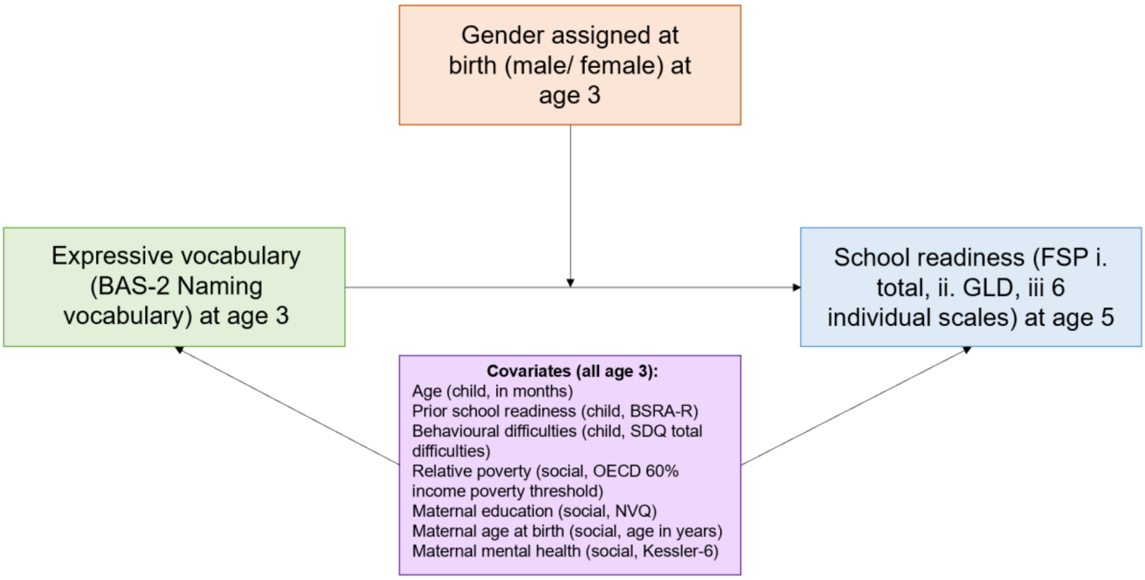
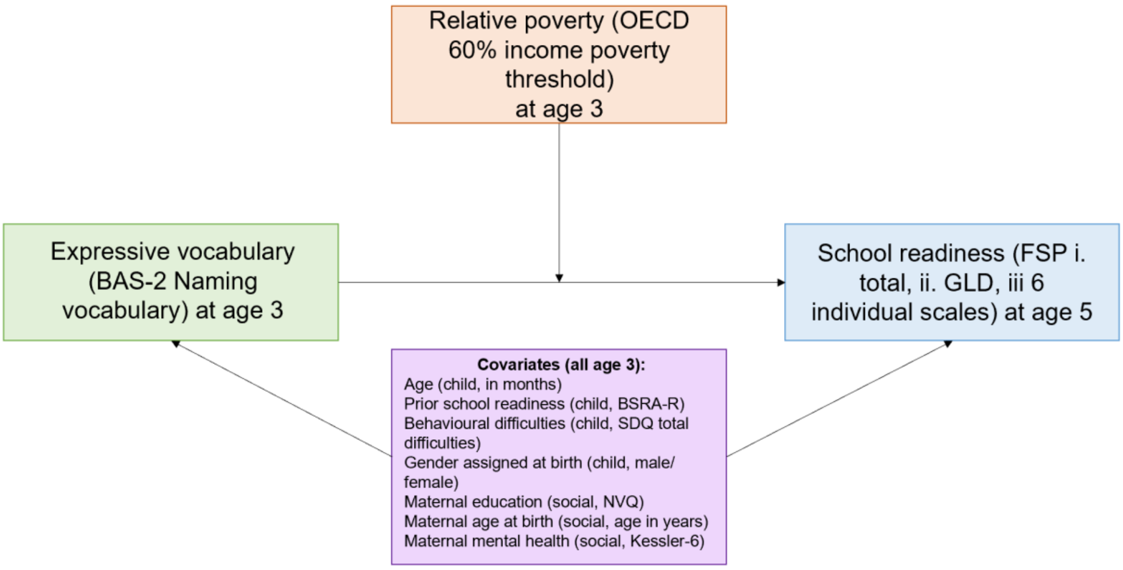


Figure 3.2. Social characteristics (relative income poverty) moderator model (model 5)



Complete case analysis

Multiple imputation was outside the scope of this study, and so complete case analyses were completed and MCS weighting used which compensates both for oversampling and, to a degree, data loss between data waves. The size of the data means that power is not an issue, but bias may still be introduced because certain populations may be more likely to be represented in the sample of children with missing data (Thabane et al., 2013). To increase the interpretability of data, children with full and partial data (i.e., had data from at least one of their moderator or covariates missing) were compared via means and t-tests/chi squares on the moderators, predictor and outcome types. Additionally, extreme outliers (via box plots and cooks distance for linear regressions Kannan & Manoj, 2015); and least likely estimates for logistic regression (Freese, 2002) were examined for BAS-2 Naming Vocabulary, FSP total and GLD data to identify outliers which may have affected the results. Extreme values were not examined for FSP scales as these each reflected part of overall school readiness. Children may have had different strengths and weaknesses in areas, and so may have been highlighted as an outlier incorrectly. As such, the total score and GLD would reflect a more holistic representation of a child's ability.

MCS weighting

The MCS had a disproportionately stratified cluster sample (e.g., oversampling of families from low SES backgrounds), meaning cases were likely to have unequal probabilities for being selected. There were also non-responses from wave 1 to 3, which needed to be adjusted for. Due to this, CLS provides a set of sample design weights which can be used to correct means, variance and non-response attrition between waves by country (Plewis, 2007). So, MCS weighting was used in adjusted analyses as it reflected the sampling design of the full English population, taking into account unequal ward selection and non-response (Schmidt & Finan, 2018). It is advised that weights used are from the latest wave (i.e. wave 3 for the current study). For the purposes of the current analysis, the single country ("weight1", "covwt1") weights were used, which weighted stratums as 1) 'advantaged' = 1.32, 2) 'disadvantaged' = 0.71, and 3) 'ethnic' = 0.24. A more detailed overview for decisions and procedure for the MCS weighting is provided in Plewis (2007). Therefore, when weighting is applied, this will not only account for missing data, but will give more weight to more affluent children due to the oversampling of children from more disadvantaged wards. In all analyses, the analyses adjusted for sample weights will be presented alongside the unadjusted values.

3.3. Results

3.3.1. Initial analyses

Initial descriptive data of in scope sample

In scope sample for study

The in-scope sample for the analyses included singleton births with complete data for BAS Naming vocabulary and FSP, living in England at wave 3 and mothers with UK occupations (n=7,012). Families or individuals which provided no data at waves 2 and 3 were excluded from analysis. The full sample selection process is presented in table 3.2. Children living in Scotland, Northern Ireland and Wales at wave 3 were excluded by default as the FSP assessment is not used in these countries.

Consideration of missing cases – Partial and full data comparison

As outlined previously, the moderation analyses were complete case analyses. Therefore, an additional analysis was completed to aid interpretation whether bias was present within the data analysed. This was completed by comparing children with complete and partial data (i.e., had data from at least one of their moderator or covariates missing). Comparisons are presented in table 3.3 between children with complete data (n=5,718) and children with partial data (n=1,294).

Children with complete data scored significantly higher in expressive vocabulary (both were within average range, but children with partial data were close to -1SD, 43.18). Children with complete data also had significantly higher foundation stage profile total and scale scores (differing by around 6 points on average in total, but no large differences for subscales). They also had a higher proportion of children achieving GLD (around 35% of children with partial data achieved GLD, while children with complete data were closer to 50% for achieving GLD). Females and children living above the relative poverty threshold were also represented more and closer to the overall selected sample proportions in children with complete data (the partial cases group had a close to 50% split for poverty, which was higher than expected as the poverty calculation should be around 40% of the sample). Furthermore, children with complete data were younger (although there was very little difference), scored better on initial school readiness (both within average range) and had lower behavioural difficulty scores (although both were within the ‘close to average’ threshold score). There was a larger proportion of children with complete data also having mothers with higher qualification levels. Children with partial data had almost a third of mothers with no qualifications compared to less than a tenth of mothers for children with complete data. Children with complete data also had mothers with lower levels of mental health difficulties (although both groups were close to low-moderate levels). There were no differences for maternal age at birth. Finally, children with complete data had a significantly higher home learning environment score. However, both groups fell around the average similar to the analysis of the entire MCS cohort by de La Rochebrochard (2012), and the original study on HLE index estimates by Melhuish et al. (2008). There were no notable differences in statistics when adjusted for sample weighting, and like the unadjusted comparisons, all but maternal age at birth were statistically significantly different.

Therefore, children with complete data demonstrated slightly better achievement and more social advantage. As such, findings from the current analyses should be considered with the caveat that completing a complete case analysis of this data means that the findings will be more representative of more socially advantaged and higher scoring children.

Table 3.2. Sample selection process

		Removal					
	N children when wave 2 and 3 datasets merged	Families > 1 cohort member (n=218 twin families = n=436 children)	Not living in England at wave 3	Not present at wave 2	No full FSP data (n=1,042)	No BAS-2 Naming Vocabulary data	Maternal education for overseas parents
n removal		-436	-7,172	-829	-1,042	-524	-225
Resulting N	17,240	16,804	9,632	8,803	7,761	7,237	7,012

Note. There were initially 19,243 families in the longitudinal dataset. 3,654 were removed for not being productive (i.e., provide data) at either wave.

Note. Triplet children families were represented as one line in the dataset. These did not have data, and so were removed before excluding children

Table 3.3. Descriptive statistics and group comparisons for all variables for children with complete versus partial data

	Unadjusted						Adjusted for sample weighting					
	Children with complete data (N = 5,718)		Children with partial data (N = 1,294)		Comparisons		Children with complete data (N = 5,718)		Children with partial data (N = 1,294)		Comparisons	
Measure	% (N)	Mean (SD)	% (N)	Mean (SD)	t / X^2	CI (95%)	Cell proportions	Mean (SE)	Cell proportions	Mean (SE)	t / X^2	CI (95%)
BAS-2 Naming Vocabulary	100 (7,012)	50.35 (10.61)	0 (0)	43.18 (13.02)	t(7010) = -20.99[a]	-7.84 to -6.50	-	50.77 (.15)	-	45.61 (.37)	t(7004) = -12.80[a]	-5.95 to 4.37
FSP total	100 (7,012)	89.12 (17.09)	0 (0)	82.43 (19.17)	t(7010) = -12.45[a]	-7.75 to -5.64	-	89.40 (.24)	-	83.16 (.59)	t(7004) = -9.78[a]	-7.50 to -5.00
FSP personal, social and emotional development	100 (7,012)	21.41 (4.04)	0 (0)	20.17 (4.43)	t(7010) = -9.81[a]	-1.49 to -0.99	-	21.47 (.06)	-	20.29 (.14)	t(7004) = -8.05[a]	-1.47 to -0.89
FSP communication, language and literacy	100 (7,012)	25.87 (6.61)	0 (0)	23.38 (7.20)	t(7010) = -12.02[a]	-2.89 to -2.08	-	25.96 (.09)	-	23.67 (.23)	t(7004) = -9.38[a]	-2.77 to -1.81

	Unadjusted						Adjusted for sample weighting					
	Children with complete data (N = 5,718)		Children with partial data (N = 1,294)		Comparisons		Children with complete data (N = 5,718)		Children with partial data (N = 1,294)		Comparisons	
Measure	% (N)	Mean (SD)	% (N)	Mean (SD)	t / X^2	CI (95%)	Cell proportions	Mean (SE)	Cell proportions	Mean (SE)	t / X^2	CI (95%)
FSP mathematical development	100 (7,012)	20.83 (4.27)	0 (0)	19.20 (4.94)	t(7010) = -12.05[a]	-1.90 to -1.37	-	20.89 (.06)	-	19.35 (.16)	t(7004) = -9.17[a]	-1.86 to -1.21
FSP Knowledge and understanding of the world	100 (7,012)	6.85 (1.51)	0 (0)	6.32 (1.74)	t(7010) = -11.25[a]	-0.63 to -0.44	-	6.88 (.02)	-	6.39 (.05)	t(7004) = -8.58[a]	-0.60 to -0.37
FSP Physical development	100 (7,012)	7.34 (1.33)	0 (0)	7.01 (1.49)	t(7010) = -7.97[a]	-0.42 to -0.25	-	7.37 (.02)	-	7.04 (.05)	t(7004) = -6.64[a]	-0.42 to -0.23
FSP Creative development	100 (7,012)	6.82 (1.44)	0 (0)	6.35 (1.54)	t(7010) = -10.36[a]	-0.55 to -0.38	-	6.84 (.02)	-	6.41 (.05)	t(7004) = -8.45[a]	-0.54 to -0.33
FSP good level of development	100 (7,012)	-	0 (0)	-	X^2(1) = 93.57[a]			-		-	X^2(1,7004) = 63.56[a]	-

	Unadjusted						Adjusted for sample weighting					
	Children with complete data (N = 5,718)		Children with partial data (N = 1,294)		Comparisons		Children with complete data (N = 5,718)		Children with partial data (N = 1,294)		Comparisons	
Measure	**% (N)**	**Mean (SD)**	**% (N)**	**Mean (SD)**	**t / X^2**	**CI (95%)**	**Cell proportions**	**Mean (SE)**	**Cell proportions**	**Mean (SE)**	**t / X^2**	**CI (95%)**
Not achieved†	49.27 (2,817)		64.14 (830)				.811		.189			
Achieved†	50.73 (2,901)		35.86 (464)				.8829		.1171			
Gender assigned at birth	100 (7,012)	-	0 (0)	-	$X^2(1) = 5.74$[b]			-		-	$X^2(1,7004) = 6.29$[c]	-
Male†	49.79 (2,847)		53.48 (692)				.8353		.1647			
Female†	50.21 (2,871)		46.52 (602)				.8578		.1422			
Poverty (OECD 60% threshold)	99.06 (6,946)	-	0.94 (66)	-	$X^2(1) = 204.79$[a]			-		-	$X^2(1,6938) = 101.32$[a]	-
Below threshold†							.783		.217			

	Unadjusted						Adjusted for sample weighting					
	Children with complete data (N = 5,718)		Children with partial data (N = 1,294)		Comparisons		Children with complete data (N = 5,718)		Children with partial data (N = 1,294)		Comparisons	
Measure	**% (N)**	**Mean (SD)**	**% (N)**	**Mean (SD)**	**t / X^2**	**CI (95%)**	**Cell proportions**	**Mean (SE)**	**Cell proportions**	**Mean (SE)**	**t / X^2**	**CI (95%)**
Above Threshold†	27.11 (1,550)		47.88 (588)				.8809		.1191			
	72.89 (4,168)		52.12 (640)									
Age (in months)	99.94 (7,008)	38.01 (2.32)	0.06 (4)	38.45 (2.94)	t(7006) = 6.20[a]	0.32 to 0.62	-	37.98 (03)	-	38.30 (.08)	t(7000) = 3.70[a]	0.15 to 0.50
Initial school readiness (BSRA-R)	92.96 (6,518)	105.36 (15.51)	7.04 (494)	94.94 (17.50)	t(6516) = -17.52[a]	-11.59 to -9.26	-	105.80 (.21)	-	97.32 (.71)	t(6510) = -11.43[a]	-9.94 to -7.03
Behaviour (SDQ)	93.52 (6,558)	9.45 (5.09)	6.48 (454)	11.24 (6.04)	t(6556) = 9.27[a]	1.41 to 2.17	-	9.35 (.07)	-	11.08 (.23)	t(6550) = 7.21[a]	1.26 to 2.20

	Unadjusted						Adjusted for sample weighting					
	Children with complete data (N = 5,718)		Children with partial data (N = 1,294)		Comparisons		Children with complete data (N = 5,718)		Children with partial data (N = 1,294)		Comparisons	
Measure	% (N)	Mean (SD)	% (N)	Mean (SD)	t / X^2	CI (95%)	Cell proportions	Mean (SE)	Cell proportions	Mean (SE)	t / X^2	CI (95%)
Maternal education (NVQ level)	99.09 (6,948)	-	0.91 (64)	-	$X^2(5)$ = 481.93[a]	-		-		-	$X^2(5,6939)$ = 63.56[a]	-
No qualifications†	8.66 (495)		31.15 (381)				.6593		.3407			
NVQ 1†	8.55 (489)		9.32 (114)				.8338		.1662			
NVQ 2†	31.39 (1,795)		26 (318)				.8655		.1345			
NVQ 3†	15.20 (869)		9.65 (118)				.9064		.0936			
NVQ 4†	32.13 (1,837)		20.69 (253)				.8904		.1096			

	Unadjusted						**Adjusted for sample weighting**					
	Children with complete data (N = 5,718)		**Children with partial data (N = 1,294)**		**Comparisons**		**Children with complete data (N = 5,718)**		**Children with partial data (N = 1,294)**		**Comparisons**	
Measure	**% (N)**	**Mean (SD)**	**% (N)**	**Mean (SD)**	**t / *X*²**	**CI (95%)**	**Cell proportions**	**Mean (SE)**	**Cell proportions**	**Mean (SE)**	**t / *X*²**	**CI (95%)**
NVQ 5†	4.07 (233)		3.19 (39)				.8654		.1346			
Maternal age at birth (years and months)	99.09 (6,948)	28.86 (5.79)	0.91 (64)	28.55 (5.94)	t(6946) = -1.68, p<.093	-0.67 to 0.05	-	28.87 (.08)	-	28.60 (.19)	t(6940) = -1.28, p<.199	-0.69 to 0.14
Maternal mental health (Kessler-6)	89.38 (6,267)	3.24 (3.63)	10.62 (745)	3.67 (3.88)	t(6265) = 2.62[b]	0.11 to 0.75	-	3.19 (.05)	-	3.54 (.17)	t(6259) = 1.98[c]	0.00 to 0.70
HLE index	99.26 (6,960)	26.75 (6.92)	0.74 (52)	23.75 (8.13)	t(6958) = -13.41[a]	-3.44 to 2.56	-	26.74 (.10)	-	24.21 (.25)	t(6952) = -9.41[a]	-3.05 to -2.00

Note. Comparisons reflect the same numbers (5,718 with complete data, 1,294 with partial data), but the N columns provide information on how many cases were present missing for each individual measure. †Percentages reflect proportions within grouping (e.g., male and female % reflects proportion for children with complete data, etc). Significant to p<.001[a], p<.01[b], p<.05[c]. Standard errors instead of standard deviations are reported for adjusted data. Cell proportions are sample proportions weighted by the MCS data.

Descriptive data of full in scope sample

Descriptive values for the full sample (N=7,012) are presented in table 3.4. On average, children scored within normal range for expressive vocabulary and foundation stage profile. As would be expected in a nationally representative dataset, gender assigned at birth proportions were very close to a split of 50% each. Less expected was a higher proportion of children living above the poverty threshold, being closer to 70%. This means the sample is more affluent on average than the general population even before complete case analysis was completed. The sample age was as expected close on average to 3 years old, and their initial school readiness was also within the normal range (standard score for average is 100). Children's behaviour was also within the 'close to average' range on average (below an SDQ total difficulty score of 13). As for the maternal characteristics, mothers were on average in their late twenties, were just outside the 'low' range of mental health difficulties (a score at or under 3), and the majority (78.69%) had obtained qualifications of at least an NVQ level 2 (at least five GCSEs A*-C or equivalent). Finally, the average Home Learning Environment was similar (but slightly higher) than the averages of Melhuish et al. (2008) and the full MCS sample analysed by de La Rochebrochard (2012) (means of 23.42 and 25.8 respectively). There were no notable differences in analyses when adjusted for sample weighting.

Extreme outliers

The presence of extreme outliers was checked before moderation analysis. There was no reason to suspect measurements were incorrect. Briefly, potential outliers were identified in very low scorers in FSP and very high scorers in BAS-2 Naming vocabulary. However, these children were retained and considered an important part of the sample, as they represented the full spectrum of scores. Further details regarding the evaluation of outliers and findings are described in appendix H.

Table 3.4. Descriptive statistics of in scope sample (all available cases) for predictor, outcome, moderators and covariates

			Unadjusted			**Adjusted for sample weights**		
Measure	**N**	**%**	**Mean**	**SD**	**CI (95%)**	**Mean**	**SE**	**CI (95%)**
BAS-2 Naming Vocabulary	7,012	100	49.03	11.44	48.76 to 49.29	49.97	.14	49.71 to 50.24
FSP total	7,012	100	87.89	17.67	87.47 to 88.30	88.44	.22	88.01 to 88.88
FSP personal, social and emotional development	7,012	100	21.18	4.14	21.09 to 21.28	21.29	.05	21.18 to 21.39
FSP communication, language and literacy	7,012	100	25.41	6.79	25.25 to 25.57	25.61	.09	25.44 to 25.78
FSP mathematical development	7,012	100	20.53	4.44	20.42 to 20.63	20.65	.06	20.54 to 20.76
FSP Knowledge and understanding of the world	7,012	100	6.75	1.57	6.72 to 6.79	6.81	.02	6.77 to 6.84
FSP Physical development	7,012	100	7.28	1.36	7.25 to 7.31	7.32	.02	7.28 to 7.35

FSP Creative development	7,012	100	6.73	1.47	6.70 to 6.77	6.78	.02	6.74 to 6.81
FSP good level of development	7,012	100	-	-	-	-	-	-
Not achieved	3,647	52.01						
Achieved	3,365	47.99						
Gender assigned at birth	7,012	100	-	-	-	-	-	-
Male	3,539	50.47						
Female	3,473	49.53						
Poverty (OECD 60% threshold)	6,946	99.06	-	-	-	-	-	-
Below threshold	2,138	30.78						
Above Threshold	4,808	69.22						
Age (in months)	7,008	99.94	38.10	2.45	38.04 to 38.15	38.03	.03	37.97 to 38.09
Initial school readiness (BSRA-R)	6,518	92.96	104.08	16.14	103.69 to 104.47	105.00	.21	104.60 to 105.40
Behaviour (SDQ total difficulties)	6,558	93.53	9.68	5.26	9.56 to 9.81	9.55	.07	9.42 to 9.68

Maternal education (NVQ level)	6,941	98.99	-	-	-	-	-	-
No qualifications	876	12.62						
NVQ 1	603	8.69						
NVQ 2	2,113	30.44						
NVQ 3	987	14.22						
NVQ 4	2,090	30.11						
NVQ 5	272	3.92						
Maternal age at birth (years and months)	6,948	99.01	28.81	5.82	28.67 to 28.94	28.83	.07	28.68 to 28.98
Maternal mental health (Kessler-6)	6,267	89.38	3.28	3.66	3.19 to 3.37	3.22	.05	3.12 to 3.31
Home learning environment (HLE index)	6,960	99.26	26.22	7.25	26.05 to 26.39	26.37	.09	26.19 to 26.55

Note. The measure ranges and thresholds are as follows: **BAS-2 Naming vocabulary:** Range 20-80, average score achieved would be 50, with +/-1SD on the score equating to +/- 10 points; **FSP total:** Range 0-117; **FSP personal, social and emotional development:** 3 subscales of 9 points each, range 0-27; **FSP communication, language and literacy:** 4 subscales of 9 points each, range 0-36; **FSP mathematical development:** 3 subscales of 9 points each, range 0-27; **FSP knowledge and understanding of the world:** Range 0-9; **FSP creative development:** Range 0-9; **FSP physical development:** Range 0-9; **FSP Good Level of Development:** achieved if a child scores an average of 6 or more across the 7 'Personal, social and emotional development' and 'Communication, language and literacy' scales; in addition to scoring 78+ points overall; **Poverty (OECD 60% threshold):** family is living in a household with net equivalent income less than 60% of the UK household; **Initial school readiness (BSRA-R):** Range: 56-149, average score achieved would be 100, with +/-1SD on the score equating to +/- 15 points; **Behaviour (SDQ difficulties):** Range 0-20, behavioural difficulties categorised as 'close to average' if between 0-13, 'slightly raised' if between 14-16, 'high' if between 17-19 and 'very high' for 20+; **Maternal education (NVQ level):** NVQ 1 = GCSE grades D-G, NVQ/SVQ/GSVQ level 1, NVQ 2 = O level/ GCSE grades A-C, trade apprenticeships, NVQ/SVQ/GSVQ level 2, NVQ 3 = A/AS/S levels, NVQ/SVQ/GSVQ level 3, NVQ 4 = first degree, diplomas in higher education, professional qualifications at degree level and nursing/other medical qualifications; NVQ 5 = higher degree; **Maternal mental health (Kessler-6):** Range 0-24, risk of psychological distress and serious mental illness categorised as low (0 to 3), moderate (4 to 6), high (7 to 12) or very high (13 to 24); **Home learning environment (HLE index):** Range 0-42, average score established by de La Rochebrochard (2012) with full MCS data was 25.8, SD = 7.39.

Moderator subgroups comparisons

Mean scores or proportions alongside comparisons tests are provided in table 3.5 (gender assigned at birth) and 3.6 (poverty). Significantly more females and children living above the poverty threshold were achieving a good level of development, and scored significantly higher in total FSP and all its scales compared to males and those living below the poverty threshold. The only exception to this was that there was no difference between genders for knowledge and understanding of the world scale. There were no notable differences in analyses when adjusted for sample weighting. As such, the highest achieving subgroups for expressive vocabulary and school readiness were females and children living above the poverty threshold.

Table 3.5. Comparisons between gender assigned at birth for predictor and outcome variables

	Unadjusted						Adjusted for sample weighting					
	Male		Female		Comparisons		Male		Female		Comparisons	
Measure	**% (N)**	**Mean (SD)**	**% (N)**	**Mean (SD)**	**t / X^2**	**CI (95%)**	**Cell proportions**	**Mean (SE)**	**Cell proportions**	**Mean (SE)**	**t / X^2**	**CI (95%)**
BAS-2 Naming Vocabulary	-	47.90 (11.35)	-	50.18 (11.42)	t(7010) = -8.39[a]	-2.81 to -1.75	-	48.73 (.19)	-	51.24 (.19)	t(7004) = -9.11[a]	-3.04 to -1.97
FSP total	-	85.45 (18.84)	-	90.37 (16.49)	t(7010) = -11.79[a]	-5.75 to -4.11	-	86.02 (.33)	-	90.90 (.30)	t(7004) = -11.03[a]	-5.75 to -4.01
FSP personal, social and emotional development	-	20.54 (4.32)	-	21.84 (3.84)	t(7010) = -13.37[a]	-1.50 to -1.11	-	20.64 (.08)	-	21.94 (.07)	t(7004) = -12.62[a]	-1.51 to -1.10
FSP communication, language and literacy	-	24.40 (7.03)	-	26.44 (6.38)	t(7010) = -12.71[a]	-2.35 to -1.72	-	24.62 (.13)	-	26.62 (.12)	t(7004) = -11.70[a]	-2.34 to -1.67
FSP mathematical development	-	20.26 (4.72)	-	20.80 (4.13)	t(7010) = -5.16[a]	-0.75 to -0.34	-	20.40 (.08)	-	20.91 (.07)	t(7004) = -4.53[a]	-0.73 to -0.29

	Unadjusted						Adjusted for sample weighting					
	Male		**Female**		**Comparisons**		**Male**		**Female**		**Comparisons**	
Measure	**% (N)**	**Mean (SD)**	**% (N)**	**Mean (SD)**	**t / *X*²**	**CI (95%)**	**Cell proportions**	**Mean (SE)**	**Cell proportions**	**Mean (SE)**	**t / *X*²**	**CI (95%)**
FSP Knowledge and understanding of the world	-	6.73 (1.62)	-	6.78 (1.51)	t(7010) = -1.34, p=.181	-0.12 to 0.02	-	6.78 (.03)	-	6.83 (.03)	t(7004) = -1.44, p=.150	-0.13 to 0.02
FSP Physical development	-	7.10 (1.45)	-	7.47 (1.25)	t(7010) = -11.49[a]	-0.43 to -0.31	-	7.12 (.03)	-	7.51 (.02)	t(7004) = -11.32[a]	-0.45 to -0.32
FSP Creative development	-	6.43 (1.53)	-	7.04 (1.34)	t(7010) = -17.94[a]	-0.68 to -0.55	-	6.47 (.03)	-	7.09 (.02)	t(7004) = -17.10[a]	-0.69 to -0.55
FSP good level of development		-		-	$X^2(1)$ = 144.39[a]	-		-		-	$X^2(1,7004)$ = 127.16[a]	-
Not achieved†	59.11 (2,092)		44.77 (1,555)				.5751		.4249			
Achieved†	40.89 (1,447)		55.23 (1,918)				.4296		.5704			

Note. Standard errors instead of standard deviations are reported for adjusted data. Significant to p<.001[a], p<.01[b], p<.05[c]. Cell proportions are sample proportions weighted by the MCS data.

Table 3.6. Comparisons between poverty for predictor and outcome variables

	Unadjusted						Adjusted for sample weighting					
	Male		Female		Comparisons		Below poverty threshold		Above poverty threshold		Comparisons	
Measure	**% (N)**	**Mean (SD)**	**% (N)**	**Mean (SD)**	**t / X^2**	**CI (95%)**	**Cell proportions**	**Mean (SE)**	**Cell proportions**	**Mean (SE)**	**t / X^2**	**CI (95%)**
BAS-2 Naming Vocabulary		43.75 (11.29)		51.41 (10.65)	t(6944) = -27.17[a]	-8.21 to -7.11		45.17 (.26)		51.87 (.16)	t(6938) = -22.24[a]	-7.29 to -6.11
FSP total		80.84 (18.56)		91.02 (16.29)	t(6944) = -22.99[a]	-11.04 to -9.31		80.89 (.44)		91.34 (.25)	t(6938) = -20.56[a]	-11.45 to -9.46
FSP personal, social and emotional development		19.85 (4.39)		21.78 (3.87)	t(6944) = -18.41[a]	-2.14 to -1.73		19.80 (.11)		21.85 (.06)	t(6938) = -17.05[a]	-2.28 to -1.81
FSP communication, language and literacy		22.64 (7.03)		26.64 (6.31)	t(6944) = -23.53[a]	-4.34 to -3.67		22.65 (.17)		26.75 (.10)	t(6938) = -20.99[a]	-4.48 to -3.71
FSP mathematical development		18.91 (4.76)		21.25 (4.09)	t(6944) = -20.88[a]	-2.56 to -2.12		18.30 (.11)		21.30 (.06)	t(6938) = -18.11[a]	-2.60 to -2.09

	Unadjusted						Adjusted for sample weighting					
	Male		**Female**		**Comparisons**		**Below poverty threshold**		**Above poverty threshold**		**Comparisons**	
Measure	**% (N)**	**Mean (SD)**	**% (N)**	**Mean (SD)**	**t / X^2**	**CI (95%)**	**Cell proportions**	**Mean (SE)**	**Cell proportions**	**Mean (SE)**	**t / X^2**	**CI (95%)**
FSP Knowledge and understanding of the world		6.23 (1.70)		6.99 (1.44)	t(6944) = -19.13[a]	-0.84 to -0.68		6.24 (.04)		7.02 (.02)	t(6938) = -16.71[a]	-0.87 to -0.69
FSP Physical development		6.92 (1.50)		7.44 (1.27)	t(6944) = -15.09[a]	-0.60 to -0.46		6.91 (.04)		7.47 (.02)	t(6938) = -13.50[a]	-0.64 to -0.47
FSP Creative development		6.30 (1.56)		6.92 (1.39)	t(6944) = -16.50[a]	-0.69 to -0.55		6.32 (.04)		6.95 (.02)	t(6938) = -14.69[a]	-0.72 to -0.55
FSP good level of development					X^2(1) = 323.76[a]	-		-			X^2(1,6936) = 292.66[a]	-
Not achieved†	68.15 (1,457)		44.78 (2,153)				.3753		.6247			
Achieved†	31.85 (681)		55.22 (2,655)				.1785		.8215			

Note. Standard errors instead of standard deviations are reported for adjusted data. Significant to p<.001[a], p<.01[b], p<.05[c]. Cell proportions are sample proportions weighted by the MCS data.

3.3.2. Moderation analysis

Testing linear and logistic regression assumptions

All figures and tables for assumption testing are presented in appendix F. For the linear regressions, the assumptions of linearity and multicollinearity between independent and outcome variables were not violated. For linearity, the scatter plot (figure 3.18) demonstrates a linear and positive trend between expressive vocabulary and foundation stage profile scores. For multicollinearity, no Pearson's correlations (table 3.15) exceeded .7 (except for the FSP outcomes with one-another, which is expected), and no VIF scores exceeded 4 (table 3.16), with a mean VIF equalling 1.76. However, when examining the scatter plot (figure 3.18) further, data points formed into a cone shape, with smaller residuals at the higher end of the scores. This suggests that the data has homoscedasticity, with smaller variance for higher values. Furthermore, the kernel density, Q-Q and P-P plots (figures 3.19, 3.20 and 3.21 respectively) suggested that residuals are not normal or expected, showing the data is more skewed at the tail ends of the distribution. The data also seems to slightly deviate from the centre of the distribution. These patterns were also seen for the six FSP subscales (syntax for testing assumptions are available here). The subscales with a maximum score of 9 had slightly more exaggerated plots (i.e., larger skew at the tail ends and centre of the distribution), but this may have been because they had a small range while the others were much larger. As for the logistic regressions, there was a binary outcome, a large sample size, and independent observations. Additionally, multicollinearity (no Spearman's correlations exceeding .7) was not violated (syntax for conducting comparisons are available here). All variables except age were linear to the log odds according to the Box-Tidwell test (table 3.17). Transforming the age variable (via taking the square root) did not have any effect in the significance (syntax for conducting comparisons are available here). When examining a scatterplot of age and total FSP (figure 3.22), the non-linearity may be due to the majority of children's ages being very similar (clustered around 3 years). However, this may still affect the robustness of log odds estimates.

Due to the evidence of homoscedasticity and non-normality of residuals in the linear regression, and non-linearity for age in the logistic regression, this would mean that the coefficients, log-odds (specifically for age) and standard errors presented in models would not be as accurate and robust. Therefore, it was suggested by Szpiro et al. (2010) to utilise a model-robust regression and a Bayesian “sandwich” estimator to correct for these issues. This correction “robust(vce)” was applied to all regressions unadjusted for sample weighting. It was not applied to adjusted analyses because the weighting already includes these corrections (STATA, 2021), and so it is not appropriate to apply a similar correction twice.

Regression results

Model variance

Foundation Stage Profile - total

Results from models 1,2 and 4 are presented in table 3.7 unadjusted and adjusted for sample weights (expressive vocabulary *gender assigned at birth) and for models 1, 3, and 5 are provided in table 3.9 (expressive vocabulary * poverty) for FSP total. When comparing this to the models also adjusted for survey weights, there were some differences. Where this occurs, differences are highlighted and discussed. All overall models were significant (all $p<.001$). As would be expected, the lowest variance explained by the model with only expressive vocabulary included (model 1; 13.65%). Gender assigned at birth as the moderator alone (model 2; 14.91%) and with covariates (model 4; 26.45%) further increased this variance. Poverty as the moderator alone (model 3; 16.08%) and with covariates (model 5; 26.30%) also further increased this variance. Therefore, this suggests that variance in school readiness total was better explained with the inclusion of the additional factors compared to expressive vocabulary alone. Variance of models slightly increased when adjusted for sample weighting, suggesting that the explanation of variables is more potent when sample is more closely representative of the English population. The models explain a small to moderate amount of variance in school readiness total scores.

Foundation Stage Profile - scales

Separate similar regressions were run for each of the six scales (Personal, social and emotional development; communication, language and literacy; mathematical development; knowledge and understanding of the world; creative development; and physical development; see tables 3.18 to 3.29 in appendix G for full statistics). Variance results were similar to the FSP total. As such, each school readiness scale was also better explained with the inclusion of the additional factors compared to expressive vocabulary alone.

Foundation Stage Profile – Good Level of Development

GLD is achieved if a child scores an average of 6 or more for each of the 7 'Personal, social and emotional development' and 'Communication, language and literacy' scales; in addition to scoring 78+ points overall. Unadjusted and adjusted for sample weight results for models 1,2 and 4 are provided in table 3.8 (expressive vocabulary and gender assigned at birth) and for models 1, 3, and 5 are provided in table 3.10 (expressive vocabulary and poverty) for GLD. Due to how the survey weights affect the model, Pseudo R^2 values could only be provided for the unadjusted models. All overall models were significant (all $p<.001$). As expected from the prior findings, the lowest variance for school readiness was explained with model 1 (7.17%), and the most variance was explained by models 4 and 5 (model 4; 13.69%; model 5; 13.68%). This suggests that GLD was better explained with the inclusion of the additional factors compared to expressive vocabulary alone. The models explain a small amount of variance in school readiness total scores.

Expressive vocabulary

Foundation Stage Profile - total

Expressive vocabulary was significant in all models (all p<.001), suggesting that it predicted school readiness. In the basic model (1), expressive vocabulary T-scores accounted for 13.65% of the variance. The coefficient indicated that for every increase of 1 point in BAS-2 Naming Vocabulary, there was between a .19 (model 5) and .62 (model 2) point increase in FSP total. In other words, for every standard deviation change in expressive vocabulary (10 points) on average, there was between a 1.90- and 6.20-point increase in school readiness total. Put in a real terms example, children scoring -1SD below the average on expressive vocabulary were falling behind on average between 1.6% and 5.30% of the raw score on total FSP scores compared to children with average expressive vocabulary scores. Adjusting any model by sample weights did not impact the significance of expressive vocabulary. Changes to coefficients demonstrated a positive trend, with increases between 0.01 to 0.10 (or between 0.10- and 1-point increase in school readiness total for +1 SD change on average) depending on the model. Therefore, when the sample is more closely representative of the English population, the effect of expressive vocabulary on children's school readiness scores is even larger. Full statistics are presented in table 3.7.

Foundation Stage Profile – scales

Like the FSP total, expressive vocabulary significantly and positively predicted outcomes in all 6 sets of models. This did not change when adjusted for sample weighting. Therefore, expressive vocabulary predicted each subscale as well as the total score. It should be noted that when examining the coefficients, these will appear to be smaller likely because the maximum scores for each subscale are smaller. Full statistics are presented in tables 3.18 to 3.29 in appendix G.

Foundation Stage Profile - Good Level of Development

Expressive vocabulary was significant in all models (all p<.001), suggesting that it predicted Good Level of Development. Unadjusted odds ratios indicated that higher expressive vocabulary scores meant children were between 1% (model 5) and 6% (models 1 and 2) more likely to achieve GLD in models. This very slightly increased (by 1%) in models 3,4 and 5 when adjusted by sample weighting. Therefore, when school outcomes are based on government expectations, expressive vocabulary consistently and positively predicts school readiness. Full statistics are presented in table 3.7.

Gender assigned at birth

Foundation Stage Profile - total

Gender assigned at birth was significant in all models (2,4,5; all p<.001), suggesting that gender assigned at birth predicts school readiness outcomes. As expected from the initial subgroup comparisons, there was a positive relationship between being female and school readiness total scores. Specifically, females scored between 2.92 (model 5) and 10.30 (model 2) points higher on the school readiness total. Put in a real terms example, males were falling behind on average between 2.5% and 8.8% on total FSP raw scores compared to females. Gender assigned at birth and expressive vocabulary significantly interacted. The interaction between expressive vocabulary and gender assigned at birth was negative (p<.001). In other words, females benefit less in their FSP scores from having higher vocabulary and males benefit more.

The coefficients for the interaction were -0.14 (model 2) and -0.13 (model 4). This meant that for every increase in average by 10 points (i.e., +1 SD) for expressive vocabulary, females benefited less than males by 1.4 points (1.2% of total score) in the model without covariates, and 1.3 points (1.1% of total score) in the model adjusted for covariates. Figures 3.3 and 3.4 illustrates a scatterplot of this interaction with the final sample numbers in both models without (n=7012, model 2) and with (n=5,718, model 4) covariates. When adjusted for sample weighting, models demonstrated no changes to significance, but the coefficients were slightly smaller, with a loss of 1.05 (model 3), 1.08 (model 4) and 2.51 (model 5) points when gender assigned at birth was an individual predictor, and -0.02 for both interactions. Therefore, when the sample is more representative of the population, females still benefited less than males, but to a smaller degree. Full statistics are presented in table 3.7.

Figure 3.3. Scatterplot of BAS-2 Naming Vocabulary and FSP split by male (orange), female (purple) and total sample (black) Model 2 (n=7,012)

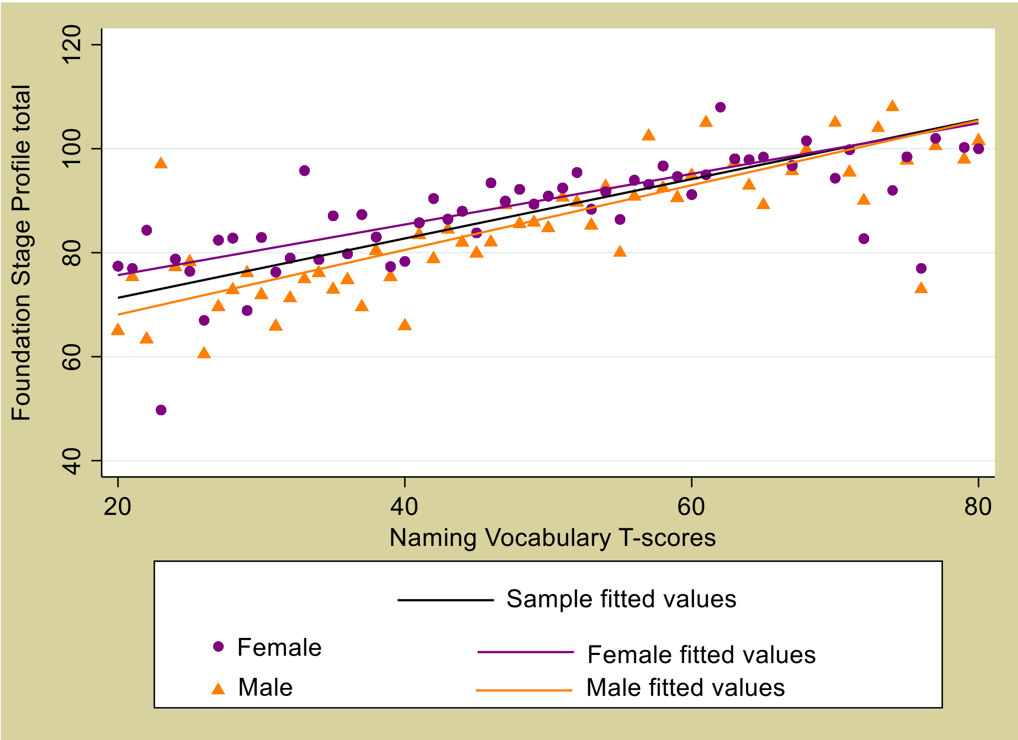


Figure 3.4. Scatterplot of BAS-2 Naming Vocabulary and FSP split by male (orange), female (purple) and total sample (black) Model 2 (n=5,718)

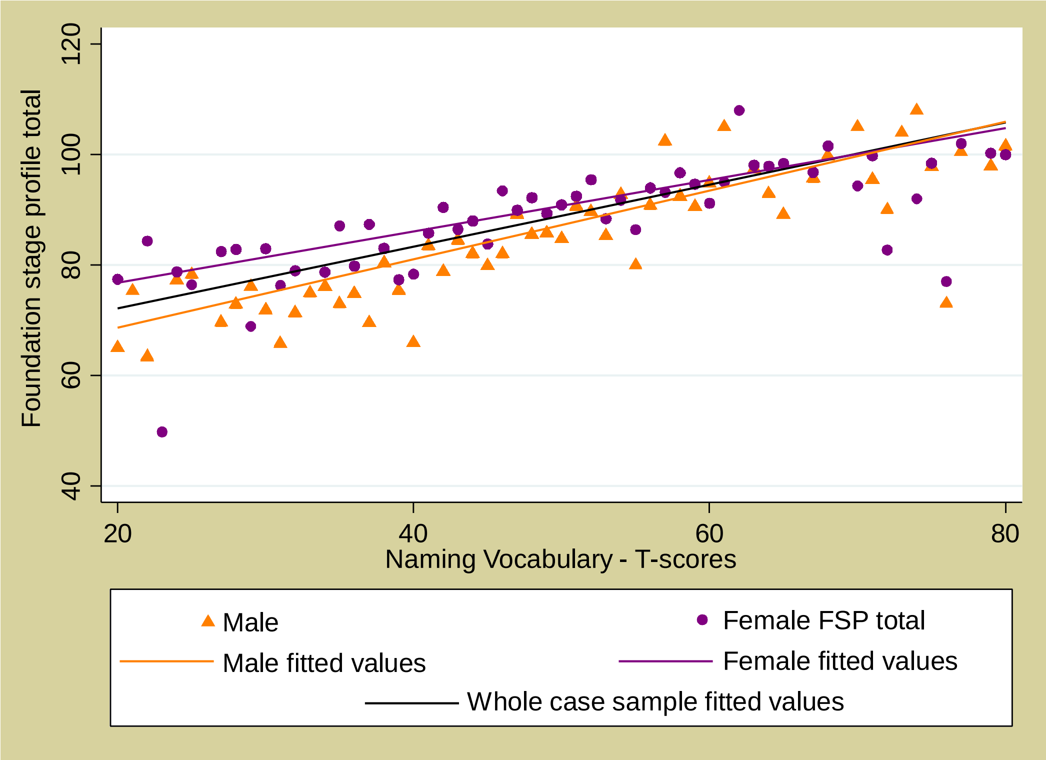


Table 3.7. Linear regression models for Foundation Stage Profile total - gender assigned at birth

	Model 1						**Model 2**						**Model 4**					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1365			**R^2 = 0.1416**			**R^2 = 0.1491**			**R^2 = 0.1523**			**R^2 = 0.2645**			**R^2 = 0.2728**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.57[a]	0.02	0.54 to 0.61	0.60[a]	0.02	0.56 to 0.64	0.62[a]	0.03	0.57 to 0.67	0.64[a]	0.03	0.59 to 0.70	0.22[a]	0.03	0.16 to 0.28	0.23[a]	0.03	0.17 to 0.30
Gender assigned at birth (female)	-	-	-	-	-	-	10.30[a]	1.85	6.67 to 13.93	9.25[a]	0.06	5.20 to 13.29	9.59[a]	2.03	5.62 to 13.57	8.51[a]	2.18	4.25 to 12.78
Interaction	-	-	-	-	-	-	-0.14[a]	0.04	-0.20 to -0.07	-0.12[b]	0.04	-0.19 to -0.04	-0.13[a]	0.04	-0.21 to -0.06	-0.11[c]	0.04	-0.19 to -0.03
Poverty (above threshold)	-	-	-	-	-	-	-	-	-	-	-	-	2.66[a]	0.55	1.58 to 3.74	2.76[a]	0.58	1.61 to 3.91

	Model 1						**Model 2**						**Model 4**					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1365			**R^2 = 0.1416**			**R^2 = 0.1491**			**R^2 = 0.1523**			**R^2 = 0.2645**			**R^2 = 0.2728**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.41[a]	0.08	-0.57 to -0.25	-0.45[a]	0.09	-0.62 to -0.27
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.33[a]	0.02	0.29 to 0.36	0.33[a]	0.02	0.29 to 0.36
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.26[a]	0.05	-0.35 to -0.17	-0.27[a]	0.05	-0.36 to -0.17
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	1.12	1.08[c]	-1.01 to 3.24	1.23	1.15	-1.02 to 3.48
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	3.03	0.90[a]	1.27 to 4.79	3.29[a]	0.95	1.43 to 5.14

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1365			**R^2 = 0.1416**			**R^2 = 0.1491**			**R^2 = 0.1523**			**R^2 = 0.2645**			**R^2 = 0.2728**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	4.18	0.96[a]	2.29 to 6.07	4.27[a]	1.01	2.29 to 6.24
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	5.78	0.93[a]	3.97 to 7.60	5.84[a]	0.97	3.93 to 7.75
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	7.02	1.20[a]	4.67 to 9.38	6.53[a]	1.27	4.05 to 9.02

	Model 1 Expressive vocabulary only N = 7,012						Model 2 Expressive vocabulary x gender assigned at birth N = 7,012						Model 4 Expressive vocabulary x gender assigned at birth with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1365			**R^2 = 0.1416**			**R^2 = 0.1491**			**R^2 = 0.1523**			**R^2 = 0.2645**			**R^2 = 0.2728**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.01	0.04	-0.07 to 0.08	0.01	0.04	-0.07 to 0.09
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	-0.05	0.06[c]	-0.17 to 0.07	-0.09	0.07	-0.22 to 0.04
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.11	0.03[a]	0.05 to 0.17	0.12[a]	0.03	0.06 to 0.18
Constant	59.91[a]	0.94	58.06 to 61.75	58.39[a]	1.04	56.35 to 60.44	55.64[a]	1.34	53.02 to 58.27	54.80[a]	1.49	51.89 to 57.71	51.43	4.07[a]	43.46 to 59.40	52.02[a]	4.22	43.74 to 60.29

Note. Significant to p<.001[a], p<.01[b], p<.05[c]

Foundation Stage Profile – scales

Gender assigned at birth showed similar patterns of predictiveness and moderation results to the total score in almost all scales, except for the knowledge and understanding of the world and the physical development subtests. For knowledge and understanding of the world, it was not a significant predictor or moderator, even when adjusted for sample weighting. For physical development, when covariates were included (model 4), gender assigned at birth also became a non-significant predictor or moderator, even when adjusted for sample weighting. However, it remained significant regardless of sample weighting as a predictor in model 5. In sum, the moderation effects for the scales scores were very similar to the overall score with the exception of knowledge and understanding of the world (table 3.24, appendix G) and physical development (table 3.26, appendix G).

Foundation Stage Profile - Good Level of Development

Gender significantly predicted school readiness in all unadjusted models (2 and 4 = $p<.01$, 5 = $p<.001$), but only remained as a significant predictor in model 5 (poverty moderation model with covariates) when adjusted for sample weighting. Unadjusted odds ratios indicated that females were between 61% (model 5, poverty moderation model with covariates) and 116% (model 4, moderation model with covariates) more likely to achieve a Good Level of Development compared to males. Odds ratios adjusted for sample weighting decreased, but still indicated that females were between 64% (model 5, poverty moderation model with covariates) and 72% (model 4, moderation model with covariates) more likely to achieve a Good Level of Development compared to males. It also did not significantly moderate the relationship between expressive vocabulary and Good Level of Development in either model (2,4) and when unadjusted or adjusted for sample weighting. Therefore, when school outcomes are based on government expectations, gender individually predicts Good Level of Development (females are achieving better), but when data is more representative of the population, the likelihood of females achieving GLD over males is lessened and mostly do not significantly differ. Furthermore, the moderation analyses suggest the proportion of both males and females achieving GLD is similar if they have higher vocabulary ability. Full statistics are presented in table 3.8.

Table 3.8. Logistic regression models for Foundation Stage Profile Good Level of Development - gender assigned at birth

	Model 1						**Model 2**						**Model 4**					
	Expressive vocabulary only N = 7,012						**Expressive vocabulary x gender assigned at birth N = 7,012**						**Expressive vocabulary x gender assigned at birth with covariates N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	Pseudo R^2 = 0.0649			**Pseudo R^2 Unavailable**			**Pseudo R^2 = 0.0753**			**Pseudo R^2 Unavailable**			**Pseudo R^2 = 0.1356**			**Pseudo R^2 Unavailable**		
Measure	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**
BAS-2 Naming Vocabulary	0.06[a] (0.00)	1.06	0.05 to 0.06	0.06[a] (0.00)	1.06	0.06 to 0.07	0.06[a] (0.00)	1.06	0.05 to 0.06	0.06 (0.00)	1.06	0.05 to 0.07	0.02[a] (0.00)	1.02	0.01 to 0.03	0.02[a] (0.00)	1.02	0.01 to 0.03
Gender assigned at birth (female)	-	-	-	-	-	-	0.74[b] (0.25)	2.09	0.26 to 1.22	0.52 (0.28)	1.68	-0.03 to 1.06	0.77[b] (0.31)	2.16	0.17 to 1.37	0.54 (0.33)	1.72	-0.11 to 1.20
Interaction	-	-	-	-	-	-	-0.01 (0.01)	1.00	-0.01 to 0.01	0.00 (0.01)	1.00	-0.01 to 0.01	-0.01 (0.01)	0.99	-0.02 to 0.01	0.00 (0.01)	1.00	-0.01 to 0.01

	Model 1						Model 2						Model 4					
	Expressive vocabulary only N = 7,012						Expressive vocabulary x gender assigned at birth N = 7,012						Expressive vocabulary x gender assigned at birth with covariates N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	Pseudo R^2 = 0.0649			Pseudo R^2 Unavailable			Pseudo R^2 = 0.0753			Pseudo R^2 Unavailable			Pseudo R^2 = 0.1356			Pseudo R^2 Unavailable		
Measure	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**
Poverty (above threshold)	-	-	-	-	-	-	-	-	-	-	-	-	0.31[a] (0.08)	1.36	0.16 to 0.46	0.35[a] (0.08)	1.42	0.19 to 0.51
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.05[a] (0.01)	0.95	-0.08 to -0.02	-0.05[a] (0.01)	0.95	-0.08 to -0.03
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.04[a] (0.00)	1.04	0.03 to 0.04	0.04[a] (0.00)	1.04	0.03 to 0.04
Behavioural	-	-	-	-	-	-	-	-	-	-	-	-	-0.02[a] (0.01)	0.98	-0.04 to -0.01	-0.02[a] (0.01)	0.98	-0.04 to -0.01

	Model 1						Model 2						Model 4					
	Expressive vocabulary only N = 7,012						Expressive vocabulary x gender assigned at birth N = 7,012						Expressive vocabulary x gender assigned at birth with covariates N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	Pseudo R^2 = 0.0649			Pseudo R^2 Unavailable			Pseudo R^2 = 0.0753			Pseudo R^2 Unavailable			Pseudo R^2 = 0.1356			Pseudo R^2 Unavailable		
Measure	Coef. (SE)	OR	95% CI	Coef. (SE)	OR	95% CI	Coef. (SE)	OR	95% CI	Coef. (SE)	OR	95% CI	Coef. (SE)	OR	95% CI	Coef. (SE)	OR	95% CI
difficulties																		
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	0.14 (0.15)	1.15	-0.16 to 0.44	0.12 (0.16)	1.12	-0.20 to 0.43
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	0.38[a] (0.12)	1.46	0.14 to 0.62	0.36[b] (0.13)	1.43	0.09 to 0.62
Maternal educatio	-	-	-	-	-	-	-	-	-	-	-	-	0.35[b] (0.14)	1.42	0.09 to 0.62	0.29[c] (0.15)	1.34	0.00 to 0.58

	Model 1 Expressive vocabulary only N = 7,012						Model 2 Expressive vocabulary x gender assigned at birth N = 7,012						Model 4 Expressive vocabulary x gender assigned at birth with covariates N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	Pseudo R^2 = 0.0649			Pseudo R^2 Unavailable			Pseudo R^2 = 0.0753			Pseudo R^2 Unavailable			Pseudo R^2 = 0.1356			Pseudo R^2 Unavailable		
Measure	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**
n (NVQ 3)																		
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	0.60[a] (0.13)	1.82	0.35 to 0.85	0.54[a] (0.14)	1.71	0.26 to 0.81
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	0.75[a] (0.20)	2.11	0.36 to 1.13	0.63[a] (0.21)	1.88	0.21 to 1.05
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.00 (0.01)	1.00	-0.01 to 0.02	0.01 (0.01)	1.01	-0.01 to 0.02

	Model 1 Expressive vocabulary only N = 7,012						Model 2 Expressive vocabulary x gender assigned at birth N = 7,012						Model 4 Expressive vocabulary x gender assigned at birth with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	Pseudo R^2 = 0.0649			**Pseudo R^2 Unavailable**			**Pseudo R^2 = 0.0753**			**Pseudo R^2 Unavailable**			**Pseudo R^2 = 0.1356**			**Pseudo R^2 Unavailable**		
Measure	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	-0.02 (0.01)	0.98	-0.03 to 0.00	-0.02[c] (0.01)	0.98	-0.04 to 0.00
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.02[a] (0.00)	1.02	0.01 to 0.03	0.02[a] (0.00)	1.02	0.01 to 0.03
Constant	-2.85[a] (0.12)	0.06	-3.09 to -2.61	-3.03[a] (0.00)	0.05	-3.30 to -2.76	-3.14[a] (0.17)	0.04	-3.48 to -2.80	(0.19)	0.04	-3.56 to -2.81	-4.03[a] (0.61)	0.02	-5.24 to -2.83	-4.01[a] (0.67)	0.02	-5.32 to -2.71

Note. Significant to p<.001[a], p<.01[b], p<.05[c]. OR = odds ratio

Poverty

Foundation Stage Profile - total

Relative poverty was a significant individual predictor in all models it was used in (2,4,5; all $p<.05$), suggesting that it predicts school readiness outcomes. As expected from the initial subgroup comparisons, there was a positive relationship between living above the poverty threshold and FSP scores. But the coefficients were markedly smaller than gender assigned at birth and expressive vocabulary. Living above the poverty threshold demonstrated an increase in score ranging between 2.66 (model 3) and 5.00 (model 5) in total school readiness. This indicated that on average children in poverty were behind by between 2.27%, and 4.27% of the raw score on total FSP score compared to their more affluent peers. Unlike gender assigned at birth, poverty was found not be a significant moderator in unadjusted analyses. The coefficients were 0.05 and -0.05 for models 3 (without covariates) and 5 (with covariates) respectively, or between 0.50 and -0.50 points (0.43% of the total score) for every +1 SD in expressive vocabulary while living above the poverty threshold. This would suggest that both children above and below the poverty threshold achieve school readiness similarly if they have higher expressive vocabulary ability. Figures 3.5 and 3.6 illustrates a scatterplot of this interaction with the final numbers in both models 3 (n=6,946) and 5 (n=5,718).

Figure 3.5. Scatterplot of BAS-2 Naming Vocabulary and FSP split by below poverty threshold (red), above poverty threshold (green) and total sample (black) (n-6,946)

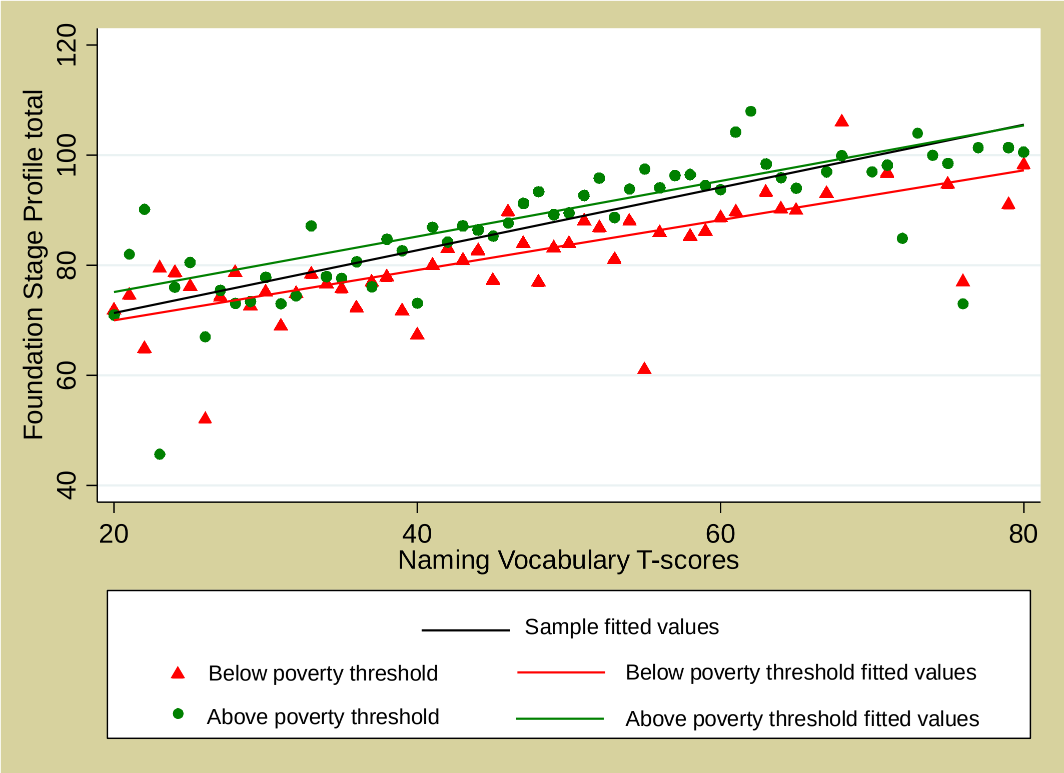
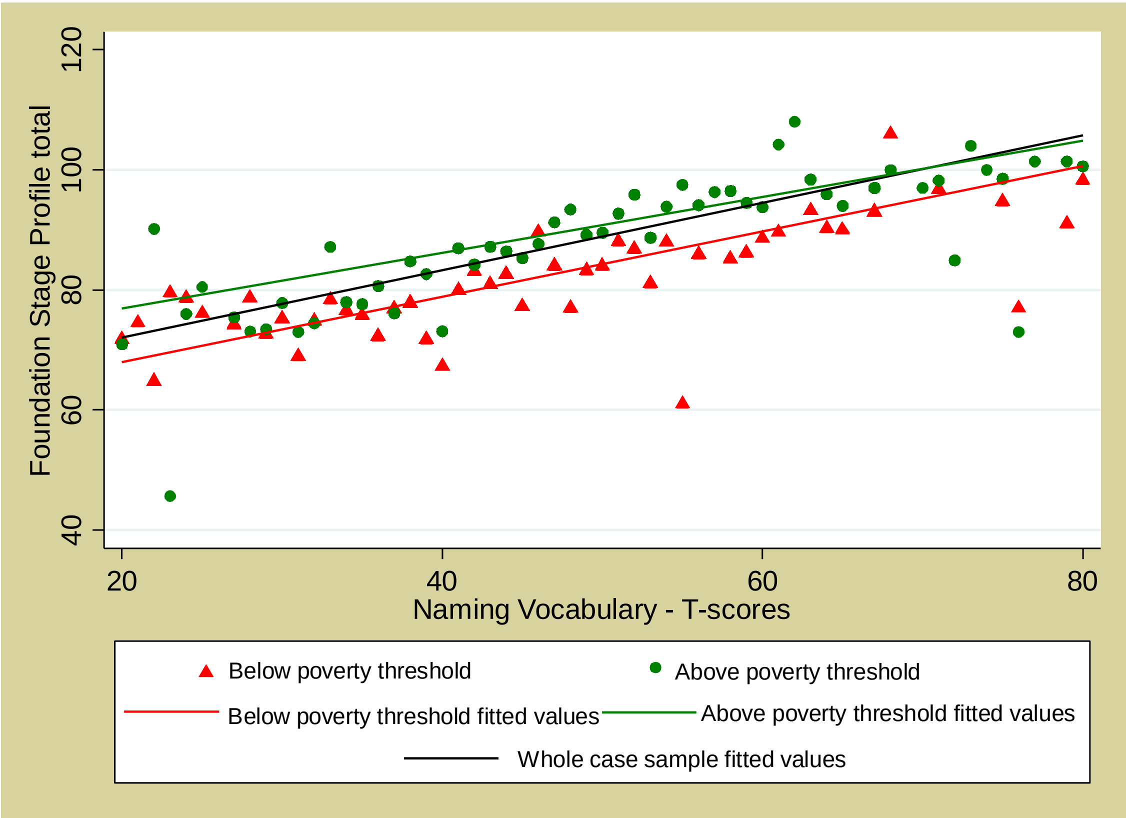


Figure 3.6. Scatterplot of BAS-2 Naming Vocabulary and FSP split by below poverty threshold (red), above poverty threshold (green) and total sample (black) (n=5,718)



However, when adjusting for sample weighting, there were some notable changes. While FSP points in model 4 (table 3.7) did not change substantially for poverty, it more than doubled in model 3 (from 4.04 to 8.96) and increased to 7.77 (from 5.00) in model 5. Therefore, children in poverty now fall behind between 6.64% and 7.66% in their total scores compared to peers, meaning there is a larger and more substantial gap when the sample is more closely representative of the English population. Additionally, while the model 3 interaction coefficient was still non-significant (but became negative, -0.04); the model 5 interaction became significant ($p<.05$) and increased from -0.05 to -0.11. So, when the sample is more closely representative of the English population, and if other child and social factors are accounted for, for every increase in of 10 points (i.e., +1SD) for expressive vocabulary for children living above the relative poverty threshold, children in poverty benefited more than those above the poverty threshold by 1.1 (0.94%) FSP total points. While indicating a small difference, when the sample is more closely representative of the English population, children not living in poverty benefit less in their FSP scores from having higher vocabulary, and children living in poverty benefit more. Full statistics are presented in table 3.9.

Table 3.9. Linear regression models for Foundation Stage Profile total – poverty

	Model 1 Expressive vocabulary only N = 7,012						**Model 3 Expressive vocabulary x poverty N = 6,946**						**Model 5 Expressive vocabulary x poverty with covariates N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1365			**R^2 = 0.1416**			**R^2 = 0.1608**			**R^2 = 0.1701**			**R^2 = 0.2630**			**R^2 = 0.2724**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.57[a]	0.02	0.54 to 0.61	0.60[a]	0.02	0.56 to 0.64	0.45[a]	0.03	0.39 to 0.52	0.55[a]	0.04	0.48 to 0.62	0.19[a]	0.04	0.11 to 0.28	0.26[a]	1.15	0.17 to 0.35
Poverty (above threshold)	-	-	-	-	-	-	4.04[c]	2.01	0.10 to 7.98	8.96[a]	2.23	4.60 to 13.32	5.00[c]	2.34	0.41 to 9.60	7.77[a]	0.95	2.83 to 12.71
Interaction	-	-	-	-	-	-	0.05	0.04	-0.03 to 0.13	-0.04	0.04	-0.13 to 0.04	-0.05	0.05	-0.14 to 0.04	-0.11[c]	1.01	-0.20 to -0.01
Gender assigned at birth (female)	-	-	-	-	-	-	-	-	-	-	-	-	2.92[a]	0.39	2.15 to 3.69	2.84[a]	0.98	2.03 to 3.65

	Model 1						Model 3						Model 5					
	Expressive vocabulary only						Expressive vocabulary x poverty						Expressive vocabulary x poverty with covariates					
	N = 7,012						N = 6,946						N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	$R^2 = 0.1365$			$R^2 = 0.1416$			$R^2 = 0.1608$			$R^2 = 0.1701$			$R^2 = 0.2630$			$R^2 = 0.2724$		
Measure	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.41[a]	0.08	-0.57 to -0.24	-0.44[a]	1.27	-0.62 to -0.27
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.33[a]	0.02	0.29 to 0.36	0.33[a]	0.04	0.29 to 0.36
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.26[a]	0.05	-0.35 to -0.18	-0.27[a]	0.07	-0.36 to -0.17
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	1.01	1.09	-1.13 to 3.14	1.10	0.03	-1.15 to 3.35
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	2.96[a]	0.90	1.19 to 4.72	3.19[a]	4.44	1.32 to 5.06

	Model 1						Model 3						Model 5					
	Expressive vocabulary only						Expressive vocabulary x poverty						Expressive vocabulary x poverty with covariates					
	N = 7,012						N = 6,946						N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	$R^2 = 0.1365$			$R^2 = 0.1416$			$R^2 = 0.1608$			$R^2 = 0.1701$			$R^2 = 0.2630$			$R^2 = 0.2724$		
Measure	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	4.10[a]	0.97	2.20 to 5.99	4.14[a]	1.15	2.15 to 6.13
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	5.75[a]	0.93	3.93 to 7.56	5.78[a]	0.95	3.87 to 7.70
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	7.00[a]	1.20	4.64 to 9.35	6.45[a]	1.01	3.97 to 8.94
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.01	0.04	-0.07 to 0.08	0.01[a]	0.98	-0.07 to 0.09
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	-0.04	0.06	-0.16 to 0.08	-0.09[a]	1.27	-0.22 to 0.04

	Model 1						**Model 3**						**Model 5**					
	Expressive vocabulary only						**Expressive vocabulary x poverty**						**Expressive vocabulary x poverty with covariates**					
	N = 7,012						**N = 6,946**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1365			**R^2 = 0.1416**			**R^2 = 0.1608**			**R^2 = 0.1701**			**R^2 = 0.2630**			**R^2 = 0.2724**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.11[a]	0.03	0.05 to 0.17	0.12[a]	0.04	0.05 to 0.18
Constant	59.91[a]	0.94	58.06 to 61.75	58.39[a]	1.04	56.35 to 60.44	61.00[a]	1.58	57.91 to 64.08	55.97[a]	1.78	52.48 to 59.45	52.87[a]	4.27	44.49 to 61.25	51.04[a]	0.07	42.33 to 59.76

Note. Significant to p<.001[a], p<.01[b], p<.05[c]

Foundation Stage Profile – scales

Poverty appeared to be an inconsistent predictor or interaction for all scales until analyses were adjusted for sample weights, and covariates were added into the model. It became a significant predictor for all subscales, and like total score became a significant moderator in model 5 for subscales except for the personal, social and emotional development (table 3.19, appendix G), and creative development scales (table 3.29, appendix G).

Foundation Stage Profile - Good Level of Development

Poverty only significantly predicted school readiness in model 4 (gender assigned at birth moderation model with covariates, $p<.001$) with and without sample weighting. However, it also became a significant predictor in model 3 (poverty moderation model without covariates) when adjusted for sample weighting. Unadjusted odds ratios indicated that children living above the poverty threshold were between 15% (model 5, moderation model with covariates) and 36% (model 4, gender assigned at birth moderation model with covariates) more likely to achieve a Good Level of Development compared to children living in poverty. Odds ratios adjusted for sample weighting increased notably, indicating that children living above the poverty threshold were between 42% (model 4, gender assigned at birth moderation model with covariates) and 80% (model 3, moderation model without covariates) more likely to achieve a Good Level of Development compared to children living in poverty. Like gender assigned at birth, it also did not significantly moderate the relationship between expressive vocabulary and Good Level of Development in either model (3,5), and when unadjusted or adjusted for sample weighting. Therefore, when school outcomes are based on government expectations, poverty is an inconsistent predictor, although children living above poverty appear to be more likely to achieve GLD than children living in poverty. This likeliness also increases when data is more representative of the population. However, both children above and below the poverty threshold achieve GLD similarly if they have higher vocabulary ability. Full statistics are presented in table 3.10.

Table 3.10. Logistic regression models for Foundation Stage Profile Good Level of Development – poverty

	Model 1						**Model 3**						**Model 5**					
	Expressive vocabulary only						**Expressive vocabulary x poverty**						**Expressive vocabulary x poverty with covariates**					
	N = 7,012						**N = 6,946**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	Pseudo R^2 = 0.0649			**Pseudo R^2 Unavailable**			**Pseudo R^2 = 0.0785**			**Pseudo R^2 Unavailable**			**Pseudo R^2 = 0.1355**			**Pseudo R^2 Unavailable**		
Measure	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**
BAS-2 Naming Vocabulary	0.06[a] (0.00)	1.06	0.05 to 0.06	0.06[a] (0.00)	1.06	0.06 to 0.07	0.04[a] (0.01)	1.04	0.03 to 0.05	0.06[a] (0.01)	1.05	0.04 to 0.06	0.01[c] (0.01)	1.01	0.00 to 0.03	0.02[a] (0.01)	1.02	0.01 to 0.04
Poverty (above threshold)	-	-	-	-	-	-	0.18 (0.26)	1.20	-0.33 to 0.70	0.59[c] (0.30)	1.80	-0.00 to 1.18	0.14 (0.34)	1.15	-0.53 to 0.81	0.46 (0.38)	1.58	-0.29 to 1.20
Interaction	-	-	-	-	-	-	0.01 (0.01)	1.01	-0.00 to 0.02	0.00 (0.01)	1.00	-0.01 to 0.02	0.00 (0.01)	1.00	-0.01 to 0.02	0.00 (0.01)	1.00	-0.02 to 0.01
Gender assigned at birth (female)	-	-	-	-	-	-	-	-	-	-	-	-	0.48[a] (0.06)	1.61	0.36 to 0.59	0.49[a] (0.06)	1.64	0.37 to 0.62

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	Pseudo R^2 = 0.0649			**Pseudo R^2 Unavailable**			**Pseudo R^2 = 0.0785**			**Pseudo R^2 Unavailable**			**Pseudo R^2 = 0.1355**			**Pseudo R^2 Unavailable**		
Measure	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.05[a] (0.01)	0.95	-0.08 to -0.02	-0.05[a] (0.01)	0.95	-0.08 to -0.03
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.04[a] (0.00)	1.04	0.03 to 0.04	0.04[a] (0.00)	1.04	0.03 to 0.04
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.02[a] (0.01)	0.98	-0.04 to -0.01	-0.02[a] (0.01)	0.98	-0.04 to -0.01
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	0.14 (0.15)	1.15	-0.16 to 0.43	0.11 (0.16)	1.12	-0.21 to 0.43
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	0.38[a] (0.12)	1.46	0.14 to 0.62	0.35[b] (0.13)	1.42	0.09 to 0.62

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	Pseudo R^2 = 0.0649			**Pseudo R^2 Unavailable**			**Pseudo R^2 = 0.0785**			**Pseudo R^2 Unavailable**			**Pseudo R^2 = 0.1355**			**Pseudo R^2 Unavailable**		
Measure	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**	**Coef. (SE)**	**OR**	**95% CI**
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	0.36[b] (0.14)	1.43	0.09 to 0.62	0.29[c] (0.15)	1.34	0.00 to 0.58
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	0.60[a] (0.13)	1.82	0.35 to 0.85	0.54[a] (0.14)	1.71	0.26 to 0.81
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	0.75[a] (0.20)	2.12	0.37 to 1.14	0.63[a] (0.21)	1.87	0.21 to 1.05

	Model 1						Model 3						Model 5					
	Expressive vocabulary only						Expressive vocabulary x poverty						Expressive vocabulary x poverty with covariates					
	N = 7,012						N = 6,946						N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	Pseudo R^2 = 0.0649			Pseudo R^2 Unavailable			Pseudo R^2 = 0.0785			Pseudo R^2 Unavailable			Pseudo R^2 = 0.1355			Pseudo R^2 Unavailable		
Measure	Coef. (SE)	OR	95% CI	Coef. (SE)	OR	95% CI	Coef. (SE)	OR	95% CI	Coef. (SE)	OR	95% CI	Coef. (SE)	OR	95% CI	Coef. (SE)	OR	95% CI
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.00 (0.01)	1.00	-0.01 to 0.02	0.01 (0.01)	1.01	-0.01 to 0.02
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	-0.01 (0.01)	0.99	-0.03 to 0.00	-0.02[c] (0.01)	0.98	-0.04 to 0.00
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.02[a] (0.00)	1.02	0.01 to 0.03	0.02[a] (0.00)	1.02	0.01 to 0.03
Constant	-2.85[a] (0.12)	0.06	-3.09 to -2.61	-3.03[a] (0.00)	0.05	-3.30 to -2.76	-2.64[a] (0.21)	0.07	-3.05 to -2.23	-3.11[a] (0.25)	0.05	-3.59 to -2.63	-3.77[a] (0.65)	0.02	-5.04 to -2.50	-4.07[a] (0.71)	0.02	-5.46 to -2.68

Note. Significant to p<.001[a], p<.01[b], p<.05[c]. OR = odds ratio

3.4. Discussion

The current study examined data from the Millennium Cohort Study and examined whether children benefit equally in their school readiness outcomes from gains in language ability, or if these benefits are moderated by additional child (gender assigned at birth) and social (poverty) factors.

3.4.1. Overall findings

Expressive vocabulary

Analyses indicated that better expressive vocabulary positively predicted FSP total score and scale scores. Furthermore, having better expressive vocabulary meant children were more likely to achieve the government benchmark of school readiness (Good Level of Development). These patterns follow those found in studies analysing the MCS data and other longitudinal research suggesting language is important for the development of school readiness skills (George et al., 2007; Hammer et al., 2017; Rodriguez & Tamis-LeMonda, 2011).

Gender assigned at birth

Gender assigned at birth was a significant predictor for total school readiness score, with males performing worse in school readiness than females. This pattern follows those found in studies analysing the MCS data and other longitudinal research (Camacho et al., 2019; Cullis & Hansen, 2008; George et al., 2007; Hobcraft & Kiernan, 2010; Mensah & Kiernan, 2010a). When analysed as a moderator, findings suggested that females benefitted less, and males benefited more from their good language in overall school readiness. This was also the case for most of the individual abilities related to school readiness. This was counter to the hypothesis suggested initially that males would gain less in their school readiness from language gains because of their lack of maturation in neurological, cognitive and socio-emotional development compared to females would affect their access to learning more generally. Why males befitted more from good language than their female peers could be because their language compensates their developmental disadvantages. This would fit with the prior research, as girls will instead be able to draw on their more globally developed skills to access learning relatively well (Adani & Cepanec, 2019; Talbot, 2020). As such, language gains are still important for gains in school readiness for females, but to a slightly lesser degree than males. If this is the case, then it may be important in future research to examine the compensatory effects of language for school readiness gains for males. However, it should be noted that the difference in the magnitude of effect between males and females is small. This suggests that while males gain more, the gap in gains is small, and both male and female children benefit from gains in language. This is encouraging as it could indicate that if children make gains in language from language interventions, then they will also likely benefit in their school readiness without a large disparity in gains based on gender subgroups.

Males and females achieved similarly in the Good Level of Development outcome. Why there is a difference between this and with total and scales scores could be due to likely converting achievement to a binary variable. Specifically, making achievement pass/fail is likely reductive as it decreases sensitivity to the variety of individual differences children have when beginning school. For example, there may be children whose skills are best in areas other than language and socio-emotional development, or children who may have just missed out from the threshold, but score relatively well in all areas. As such, findings based on the government benchmark of school readiness may not best reflect how child and social factors moderate the effect of language gains on school readiness gains in the population where developmental profiles are more heterogeneous than is allowed from their measure. This may reflect a wider issue which indicates that the government's benchmarks may not be taking into account children's individual needs when entering school; and instead places onus on the child to be ready for the curriculum and educational system as suggested by those critical of school readiness measures (Kay, 2018; Pretti-Frontczak, 2014; Roberts-Holmes, 2019).

There were two areas where gender assigned at birth did not predict or moderate effects. The first was knowledge and understanding of the world. No differences based on gender may have occurred because children are encouraged to learn about more general topics (e.g., people, places and nature) via their own experiences and interests. As such, the developmental maturity advantage that females have may be less relevant because children can engage with these learning goals in a more individualised manner rather than having to follow a more strictly prescribed lesson or task.

The second scale was physical development. There is evidence to suggest males demonstrate more development in physical skills than females in early childhood (e.g., motor skills and movement; Junaid & Fellowes, 2006; Masnjak, 2017). However, the early learning goals also included items requiring language skills more general engagement (e.g., listening to instructions, travelling around PE equipment well, understanding health and fitness). As such, early learning goals are not purely based on physical development. Instead, it also assesses skills which may require good language, communication and social skills. Therefore, females may be scoring similarly to males because they can rely on their developmental advantages in other domains to achieve early learning goals in this scale.

Poverty

Poverty was a significant predictor for total school readiness score, with children living in poverty performing worse in school readiness than their more affluent peers. This pattern follows those found in studies analysing the MCS data and other longitudinal research (Blanden & Machin, 2010; Camacho et al., 2019; Cullis & Hansen, 2008; Dex, 2008; Dickerson & Popli, 2016; Feinstein, 2003; George et al., 2007; Isaacs, 2012; Kiernan & Mensah, 2009; Law et al., 2011; Locke et al., 2002; Mensah & Kiernan, 2010a, 2010b). When analysed as a moderator, results indicated that children living above the poverty threshold benefitted less, and those living in poverty benefit more in their school readiness from good oral language. This was also the case for most of the individual abilities related to school readiness. This was counter to the hypothesis suggested initially that children living in poverty are more likely to have barriers to educational resources and enriching learning experiences which would reduce their development in multiple components of school readiness; and in turn would affect engagement with more formal learning activities. Why children living in poverty befitted more from good language than their more affluent peers could be because their language compensates for their developmental disadvantages. This would fit with the prior research, as more affluent children could instead draw on their additional resources and experiences of engaging with formal learning activities (Duncan et al., 2014; Hobcraft & Kiernan, 2010; Illøkken et al., 2021; Mollborn et al., 2014). As such, language gains are still important for gains in school readiness for children living above the poverty threshold, but to a slightly lesser degree than children living in poverty. If this is the case, then it may be important in future research to examine the compensatory effects of language for school readiness gains for children living in poverty. However, similar to gender assigned at birth it should be noted that the magnitude of effect between children living above and below the poverty threshold is small. This suggests that while children living in poverty gain more in school readiness, the gap in gains is small, and both children living above and below the poverty threshold benefit from gains in language. This is encouraging as it could indicate that if children make gains in language from language interventions, then they will also likely benefit in their school readiness without a large disparity in gains based on poverty subgroups. Also similar to gender assigned at birth, poverty did not moderate Good Level of Development. This is for likely similar reasons stated for gender.

Three areas where poverty did not predict or moderate effects were personal, social and emotional development; knowledge and understanding of the world, and creative development. As mentioned for gender assigned at birth, knowledge and understanding of the world reflects because children are encouraged to learn about more general topics (e.g., people, places and nature) via their own experiences and interests. It is also likely the resources to engage in these activities are provided by the school. Therefore, to engage in these tasks will not necessarily be reliant on family resources or the developmental disparities created from such. Like knowledge and understanding of the world, creative development assesses activities where resources are similarly available to children, and may similarly alleviate potential resource and developmental differences between children in poverty and more affluent peers. For personal, social and emotional development, this was an unexpected finding as children living in poverty are more likely to have worse socio-emotional outcomes (Lee & Zhang, 2021). However, another study analysing MCS data found that strong language dampened the effects of social deprivation on behavioural problems (Flouri et al., 2012). While this is examining language as the moderator rather than the predictor, the current is able to demonstrate that when the other way around, poverty has little effect on the relationship between language and socio-emotional skills. As such, stronger language may be a protective factor for socio-emotional risks seen in children living in poverty.

3.4.2. Strengths and limitations

The secondary data analysis to the author's knowledge is the first of its kind to examine moderation effects of gender assigned at birth and poverty on the relationship between oral language and school readiness in a population representative cohort. The final analysis included over 5,500 preschool aged children who all completed the same measures for oral language, child and social factors and school readiness. Analyses applied weighting, which took into account attrition levels and underrepresented populations, and so allowed for confidence that findings are representative of the English population. In addition, a thorough examination and selection process for the best quality measures was conducted. Therefore, the study was able to provide more representative, valid and robust findings than in the systematic review, and conclusions could be made with more confidence. The examination of extreme outliers also concluded bias was unlikely to be introduced from highly exceptional or erroneous cases.

Although multiple imputation might have offered some advantages, this was not possible within the time and resource constraints of this thesis. Instead, a complete case analysis with an in-depth examination of missing cases was conducted, and produced a good examination of potential biases in the data. This analysis indicated that children with partial data (and therefore not included in final moderation models) were more likely to be from disadvantaged subgroups or be lower scorers for oral language and school readiness. However, adjusting for the sample weights takes into account the sampling and loss of the dataset, which would partially ameliorate this issue. As such, applying the conclusions of these findings needs to be completed with the caveat that social disadvantage effects may be less reliably estimated.

The only oral language measure utilised was expressive vocabulary. Other aspects of expressive, as well as receptive language need to be considered if findings are to be applied to oral language more generally. However, vocabulary is thought to be a good indicator of broader language development up to at least 6 years old (Bishop et al., 2017; Tomblin & Zhang, 2006). Furthermore, previous longitudinal research has yielded significant insights into the long-term impacts of vocabulary. For example, Westrupp et al. (2020) found that preschool vocabulary difficulties were related to poorer socio-emotional and academic functioning across development and into adolescence. Furthermore, Willoughby (2020) found that poorer vocabulary in early childhood predicted poorer educational attainment, cognitive development, socio-emotional outcomes through to adulthood. (Coloma et al., 2020) also found that vocabulary was a robust predictor of literacy attainment during primary school. As such, this measure is likely a good representative of language development when applying the current study's findings.

The BAS-2 Naming Vocabulary and initial school readiness were only completed to children who could speak English (and Welsh for BAS-2) well enough to complete the tests, and parents and assessors could also decide whether children may not be able to complete the assessments due to additional needs. Additionally, for reasons outlined in the variable selection process (section 3.2.3 and appendix E), children whose mothers had overseas qualifications were removed from analysis as it was unclear how to categorise. Therefore, the data may also be likely to underrepresent children with a multilingual background or with special educational needs. As such, application of these findings needs to be completed with the caveat that the prevalence of these two groups may be underestimated in the analyses.

3.4.3. Conclusions

This study assessed the assumption that children will benefit equally in school readiness outcomes from gains in language ability. It was found that children assigned male at birth and living in poverty demonstrate benefit more in school readiness outcomes from language gains. While children may be initially disadvantaged from being in 'at risk' groups, having good oral language could be a protective factor which may ameliorate the effects of these developmental and social disadvantages. However, children in the 'at risk' groups may benefit more from language gains as they have less developmental and social advantages to draw on. The government benchmark of school readiness (Good Level of Development) is likely not sensitive enough to the individual differences of children, and so may not be able to pick up on whether school readiness benefits from gains in language are moderated by additional child and social factors.

There are some components of school readiness where children from different gender and poverty subgroups benefitted equally from gains in their language. For knowledge and understanding of the world, and creative development, this may because completing these activities are less reliant on the developmental and social advantages female children and children living above the poverty threshold have. For physical development, this may be because it has tasks associated with completing the skill which do not encapsulate the skill itself (e.g., following instructions). This results in allowing children to compensate for potential developmental disadvantages the skill (e.g., fine motor skills) if they have developmental advantages in learning goals related to completing the skill (e.g., socio-emotional skills).

Overall, the findings are very encouraging when considering language interventions, as it demonstrates that addressing language difficulties for children in 'at risk' groups can help them become as school ready as their more advantaged peers. This also supports the need to ensure all children in 'at risk' groups have access to early preventative language intervention. Not only because having good oral language is clearly beneficial for these children; but also because they are likely to have less advantages to draw from compared to children in the more advantaged groups.

3.5. Appendices

3.5.1. Appendix D. Dataset and variable set-up

Selection of datasets

Below are the datasets selected and the variables taken from each. Specific variable locations are provided in the MCS data dictionaries.

Table 3.11. Datasets used for variables

Dataset	**Individuals in dataset**
Longitudinal family file (*)	All families noted in the MCS ever – one line per family

Dataset	Individuals in dataset
Geographically linked data	All families noted for specific wave – one line per family
Family derived variables	All families noted for specific wave – one line per family
Household grid	All individuals for that wave – one line per person
Cohort member cognitive assessment	All cohort members noted for specific wave – one line per cohort member
Cohort member derived variables	All cohort members noted for specific wave – one line per cohort member
Parent interview about cohort member	All main and partner responders who answered questions about cohort member – 1 or 2 lines per responder depending on how many cohort members there were
Parent derived variables	All main and partner responders who answered questions about themselves – 1 line per main and partner responder
Foundation Stage profile (available at wave 3 only)	All cohort members noted for specific wave – one line per cohort member

Note: *This was a dataset which had overview data for all families in the dataset from waves 1-6.

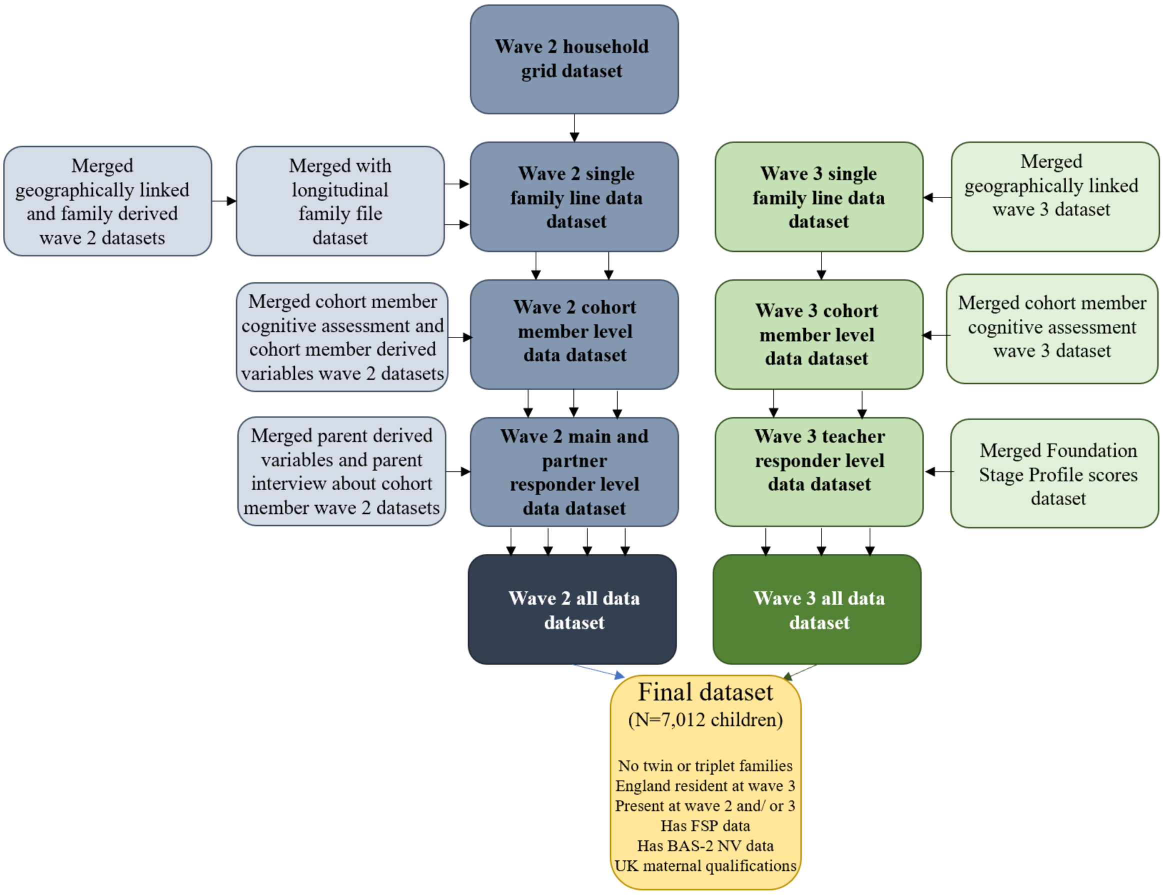
Merging and data cleaning

Due to the data required being present in multiple datasets, a planned set of merges was needed within and between the MCS waves. All merging syntax in STATA was in part developed by guidance documents from the CLS and STATA (Gould & Emeritus, 2011a, 2011b; Vilma & Johnson, 2020). As a brief overview, a modified version of each dataset was first created to only include relevant variables (e.g., analysis variables, weighting variables, identity variables) to aid with the manageability of the data. Some variables were also created at this stage for later dataset exploration and data removals (e.g., a variable which combined the eligibility and actual response of people at waves 2 and 3). One additional preparation before merging was for the datasets where responders were asked about the cohort member children. This was formatted so that there was an individual line for each child per responder. For example, if a family had two cohort member children, there were two main responder lines, and two partner responder lines if applicable. As this would cause issues for merging (i.e. there would be more than one line per person), the datasets that did this were wide-reformatted so that multiple children were placed on a single line per responder (although these would be later removed).

Datasets with the same type of data level (family, responder, cohort member) were then merged together first, followed by merging within their wave and to the wave's household grid dataset. At each merge, checks for non-merged cases, duplications or incorrectly aligned data were conducted to ensure merges was correct, and investigations were made if any of these issues occurred. Incorrectly aligned data and duplications were not an issue, and non-merged cases were usually due to individuals not being in both datasets. If any clear errors were found in the dataset were to prevent proper merging, these were corrected. No major errors were found, and only a small number individual-level errors from CLS data inputting were found. For example, one person was given two different person numbers in different datasets, although it was clear they were the same person when examining their data. Once this had been done separately for waves 2 and 3, both of these complete datasets were merged together to form the final dataset needed for analysis, and the same checks were completed. Again, no major issues were found. Figure 3.7 illustrates the merging procedure.

Data cleaning removed anyone but the cohort members in the analysis, as they were the point of interest. Before this, the main/mother-reported variables needed to be put on the same line as the child. A separate dataset with mains at wave 2 was created with the main/mother level variables and then re-merged (via their family MCS id) to the main dataset so that the data for these variables was on every family member's line. Family level variables (OECD 60% median, household English language status) were provided on all lines, so no additional steps were needed to place on the cohort data line. Once the merges were completed, misaligned data was checked for and none were found. Then, children were removed according to the exclusionary criteria. At the end of the data cleaning process, 7,012 cohort members from England were included in the analysis. See table 3.2 for the removal process for cohort children.

Figure 3.7. Merging process for datasets



Recoding variables for analysis

All variables at wave 2 were re-coded to also include a non-response code for cohort children who were not present at time 2, and if it was a cohort-level variable, added an additional code for non-cohort members. This is so they were not counted as missing in the same way that any actual missing data from included individuals who should have had data, and made it easy for them to be removed. The exception to adding missing codes was gender assigned at birth, which was matched up the reported data at wave 3, and added to included children with this missing data. This was done because of the unlikeliness for changes in gender identity to be noted between age 3 and 5. Indeed, there were no errors or changes in the data between waves for any child in the final dataset when checked. Once the final dataset was ready, and children were removed for the above reasons, all missing codes left for variables were condensed into one missing code (i.e. '.' In STATA) for ease of analyses (minus numbers, which missing codes used in the MCS datasets, were analysed with, were counted as actual number in the analyses).

The outcome variable was kept in its original total form, but was also transformed into a binary 'Good level of development' (GLD) criteria (yes/no). For the purposes of the current study, age data were transformed from days into months (but is still continuous in nature). This was to help with the interpretability of the results, because this is more commonly used in research, and can be easily translated into years for readers if required. The months were 30 days as this would add up the closest to 365 (29 = 348; 30 = 360; 31 =372). The exact transformation could not be done, as the full birth date was not provided in the dataset. To ensure that this transformation did not effect findings, both the month and days variables were placed into correlation matrices with BAS-2 Naming Vocabulary and FSP, and showed very similar results (months r=-.0246 and days r=-.0212 for BAS-2 Naming Vocabulary, and months r=-.0640 and days r=-.0647 for FSP). This indicated that age in months is a valid indicator of children's age in the sample while also being easy to interpret. Although the NVQ categories are clearly ranked, overseas qualifications did not fit with this (as qualifications could have been at any equivalent level in the respective country). Due to this and because the variable should be able to be easily interpreted by the 'amount' of education, these responses were excluded from analysis (n=225). Table 3.12 provides a summary of changes made to each variable.

Table 3.12. Recoding of variables used in analysis

Variable	Variable changes (that are not adding missing codes)	Variable Type
BAS-2 NV (predictor)	None	Continuous
FSP (outcome)	Changed to also have a variable which denotes if the child achieved the 'Good level of development' threshold outlined by the Department of Education (2010)	Continuous and Binary
Gender assigned at birth (moderator)	Added in any missing data at wave 2 from data at wave 3	Binary
OECD 60% median	None	Binary
Age (covariate)	Changed from days into (30 day) months	Continuous
BSRA-R (covariate)	None	Continuous
SDQ total behavioural difficulties (covariate)	None	Continuous
Maternal Education (covariate)	Re-coded so that no qualifications were ranked below No qualifications. Removed the overseas responses.	Ranked categorical
HLE index (covariate)	Transformed each of the item responses into a total score, the syntax was adapted from de La Rochebrochard (2012)	Continuous

Variable	Variable changes (that are not adding missing codes)	Variable Type
Maternal age (covariate)	Age was provided in years for the mother at the time of interview, so made a variable which was: parent age at wave 2 - child age (months)	Continuous
Maternal mental health K-6 total (covariate)	None	Continuous

3.5.2. Appendix E. Predictor, outcome, moderator and variable descriptions and selection considerations

The following section described the measures for each variable in more detail. Furthermore, the selection process of the moderator and covariate variables based on theoretical and methodological reasoning is outlined. The theoretical basis to include as potential moderators was highlighted in the background section, but further theoretical considerations are included here to decide what should be a moderator, covariate and excluded. While the predictor and outcome were chosen in part due to their availability, their methodological features were generally robust and outlined here.

Expressive vocabulary: British Ability Scales 2 Naming Vocabulary subtest

As highlighted in the variable selection process section (3.2.3), British Ability Scales 2 naming vocabulary was the predictor used. Reasons for its selection based on theory are described in the chapter, and so will not be outlined here. Instead, selecting the measure type and the assessment of its methodological qualities are outlined below.

Selecting score type

In the MCS, raw, ability, T-score and percentile score types are available. To obtain the best level of interpretability, these types were compared and one was chosen for the final analysis. This was aided by Connelley (2013), who completed an overview of all the cognitive and psychological measures of the MCS at wave 2. Standardised scores were eventually chosen. First, the measure in the MCS attempts to preserve the child's self-confidence, keep them motivated, and with consideration to their age and ability. Therefore, different starting and stopping times were used. Due to this, not all children answered the same items, meaning that raw scores may not represent a child's level of ability. Although ability scores take into account to difficulty of items the child completed, this is not a continuous measure and does not take into account age, meaning that some data would be lost. Additionally, children are not interviewed at the exact same age in the MCS, and so this would potentially give a disadvantage to some children. Percentile scores were also not chosen because although they account for age, they are not continuous but are instead categorical. This would mean losing data and therefore sensitivity. As such, T-scores were picked because they addressed the issues of the other three types; it accounts for age and difficulty of items completed, as well as being continuous.

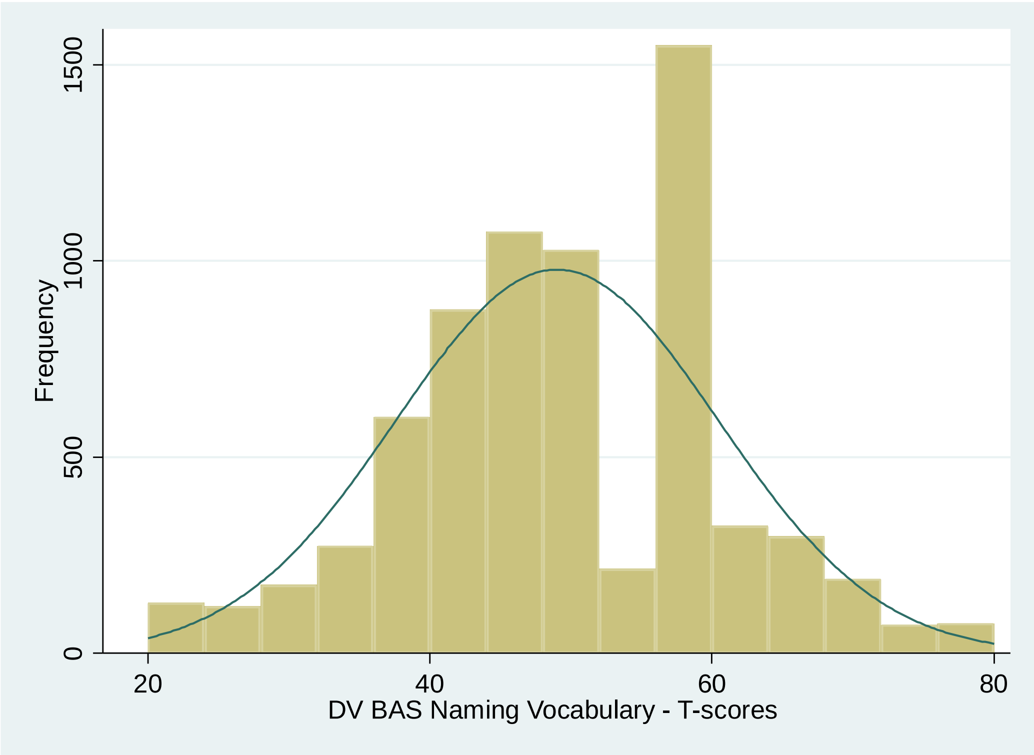
Methodological assessment

The BAS-2 is a standardised measure which has a high test-retest reliability and is normed to be representative of UK population. Selections of samples were based on type of school attended, region of residence, free school meal entitlement, gender, parental education and ethnicity (Connelly, 2013). It is also suitable for children aged from two and a half years until late adolescence (17;11 years). The type of measurement (naming pictures of objects) is a standard way in which expressive vocabulary is examined (e.g. Peabody Picture Vocabulary Test, Dunn & Dunn, 2007; Expressive Vocabulary Test, Williams, 1997-2007; Kilifi Naming Test, Kitsao-Wekulo et al., 2019). It has a high level of construct validity and test-retest reliability (Connelley, 2013).

However, an issue with the measure is that the age bins are calculated in 3-month intervals. Connelley (2013) notes this as a potential issue because children aged 3 will still be varying widely in their cognitive development. This is resolved by utilising the age variable (detailed further in 6.2.6) as a covariate in the analysis to account for this. Furthermore, due to some children's English ability or cognitive or physical disabilities, they may have been excluded from assessment. Parents were consulted on if a child had a physical or mental disability which would make them unable to complete the assessment, and if so, they did not take part; and were also not assessed if they could not speak English or Welsh (Centre for Longitudinal Studies & GfK NOP Social Research, 2006a). As such, the BAS-2 Naming Vocabulary scores may not be representative of children more likely to have lower scores, leading to a bias in the data and 'pulling up' the average score of the sample. When examining the histogram (figure 3.8) of scores alongside the skew (0.11) and kurtosis (3.04), this suggested that the data did indeed have a heavier skew to the higher end of scores. There seems to be a large number of children scoring just below 60 (i.e. +1SD above average) but a large dip in the bin before that. It is unclear why this is, and is not mentioned in the CLS documentation.

However, when examining the descriptive data for BAS-2 Naming Vocabulary of the in-scope sample, the full range (20-80) of scores were present, and the (unadjusted) mean was 49.03(11.44). This suggests that on average, the children are achieving the almost exact mean of the measure, which would suggest a representative sample. Although this does alleviate some concerns that data would be skewed to a specific level of ability, it is still important to note it may not represent some subgroups of children (with conditions or not English speaking), and there is an unusual data spike in the data.

Figure 3.8. Histogram of BAS-2 Naming Vocabulary scores (N=7,012)



School readiness: Foundation Stage Profile (government data)

As highlighted in the variable selection process section (3.2.3), and moderation analysis section (3.2.4), Foundation Stage Profile was the outcome used. Reasons for its selection based on theory are described in the chapter, and so will not be outlined here. Instead, selecting the measure type and the assessment of its methodological qualities are outlined below.

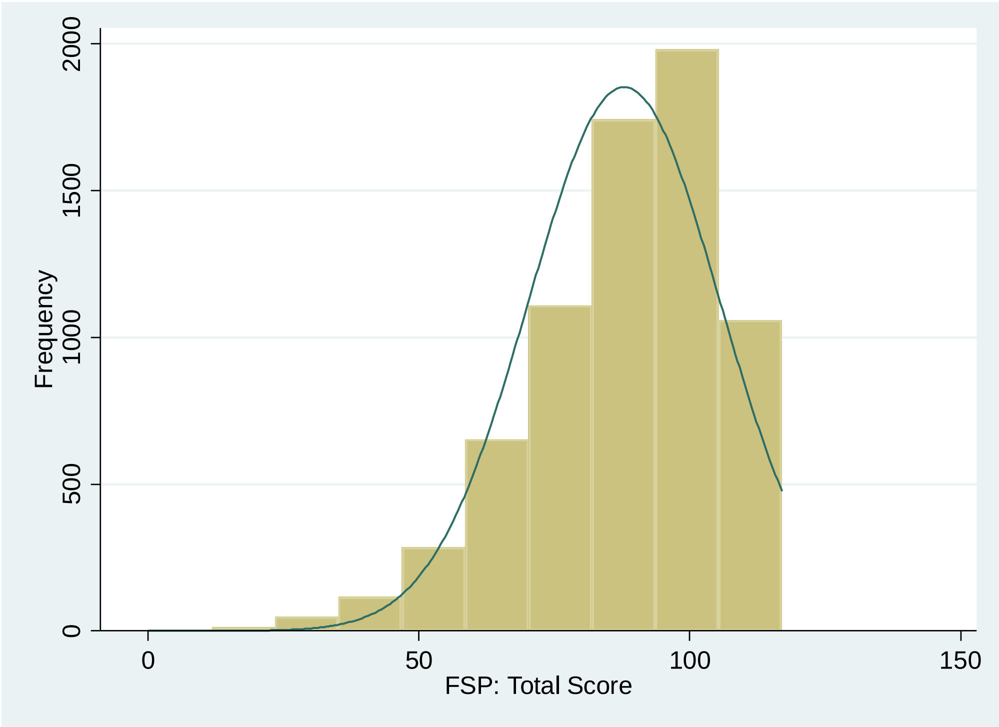
Selecting school readiness measure

At age 5 in the MCS, there were two sets of school readiness outcomes. These were the Foundational Stage Profile (FSP) for those living in England, and the Devolved Administration Teacher Survey/ Celtic Country Teacher Survey (DATS/CCTS) in Wales, Scotland and Northern Ireland. For the feasibility of the study (time and scope), DATS/CCTS was not used. To be able to complete the FSP, teachers had to undertake training to use it, and the measure is moderated at both local authority and nationally (Johnson, 2008). It is also expected to be built up throughout the year and based on cumulative evidence (Department for Education and Skills, 2003). Although the DATS/CCTS attempts to measure similar areas, teachers completing these questionnaires are not subject to the same requirements, and the implication is that the assessments are completed in one sitting as opposed to over the year. Therefore, the reliability and validity rates may not be the same, and so cannot be merged or compared. Cohort children's FSP data was obtained by the Department for Education and Skills (now Department for Children, Schools and Families) and around 95% of the full sample was successfully matched by researchers of the MCS (Hansen & Jones, 2008; Hopkin et al., 2009).

Methodological assessment

The FSP is made up of scales which include skills reflecting the level of challenge of level 1 and level 2b of the national curriculum at the time (Hansen & Jones, 2008). The Foundation Stage Profile is also employed by the government, and required to be completed by all teachers for all children. As such, the measure demonstrates a high level of validity, and its standardised nature and large-scale application means its reliability and representativeness is also of a high standard. The full MCS's cohort mean FSP total score was 87.7 (n=8,563) was similar to Nationally reported scores (Department for Education and Skills, in Hansen & Jones, 2008), meaning that FSP scores are likely representative of the whole English population. When examining the in-scope sample, the mean(SD) was 87.89 (17.67), meaning that the scores were almost exactly the same. There does seem to be a bias towards higher scores in the FSP total (skew -0.81; kurtosis 3.62, see figure 3.9 for histogram), but in the context of the measure, if the average is around 87/88 out of 117 (when the lowest score is 0) then the bell curve will have a heavier end.

Figure 3.9. Histogram of FSP total scores (n=7,012)



Note. FSP total ranged from 0-117, each subscale score had a maximum of 9, and only students with data for all subscales were included.

Gender assigned at birth: Main responder reported (CAPI)

Measure description

Gender assigned at birth is traditionally characterised in developmental research with binary ‘male’ and ‘female’ categories, also seen in the MCS. Without a complex discussion of gender identity, the variable is labelled ‘gender assigned at birth’ to expressively delineate this is what the child is identified at this age and likely since birth. The gender assigned at birth data is collected via the main respondent in the CAPI questionnaire.

Theoretical considerations

Theoretically, gender assigned at birth has good links to language and school readiness, and current research indicates it directly affect school readiness (as seen in section 3.1.2). It has also yet to be explored in high-quality moderation analysis (as noted especially in the systematic review). Therefore, it had good potential to examined as a moderator.

Methodological considerations

Gender assigned at birth was also strong methodologically, as it demonstrated high validity, reliability, representativeness and data quality. Except for human error, acquiring this data appears to have little risk of being unreliably obtained, and the question asking for gender is relatively straightforward in the CAPI questionnaire. Validity-wise, assigned gender will likely be impactful at this age, as children in early childhood are still under strong subjection to gender binarism (e.g. Callahan & Nicholas, 2019), and are demonstrating developmental differences (as outlined in section 3.1.2; Adani & Cepanec, 2019; Lovas, 2011; Maguire et al., 2016; Masnjak, 2017; Talbot, 2020; Unterrainer et al., 2013). Male and female proportions are relatively equal, with an almost 50:50 proportion (*n*=3,539 (50.47%) male, *n*=3,473 (49.53%) female). As such, it is expected that the variable would be relatively applicable to the general population. The level of response is 100% (after variable set-up). Therefore, it had good potential to examined as a moderator.

Relative income poverty: OECD 60% poverty threshold (family data)

Measure description and selection

The measure was developed by the Organisation for Economic Co-operation and Development (OECD) who are an international organisation creating operationalised equivalised income scales and conducting global research examining household incomes of multiple countries including the UK (Organisation for Economic Co-operation and Development, 2021). It notes if a family is living in a household with net equivalent income less than 60% of the median UK household,they are living in relative poverty. The MCS banded income was by CLS into an OECD equivalised score. The equivalisation definition details for OECD measures at and before 2011 is detailed in their OECD documentation (Organisation for Economic Co-operation and Development, 2012), but the calculation for equivalisation is that all incomes are adjusted by the square root of the household size. There were eighteen MCS income which ranged from £0-1700 to £85000+. Rather than the 50% equivalised median used by the OECD, the MCS modified this to be below 60%, and also had slightly adjusted equivalisation rates. Documentation detailing this adaptation did not explicitly state why this modification was made. However, the OECD documentation states two definitions of poverty, relative and absolute. In the relative definition, the OECD (2012) states “Two relative poverty thresholds are used: the first one is set at 50% of the median equivalised disposable income of the entire population, the second one is set at 60% of that income” (p2). It can therefore be implied that the MCS chose the latter of the two distinctions.

There were many variables within the MCS which related to income, and included both gross and net versions. However, the OECD measure of relative poverty was ultimately chosen as a moderator due to the quality of the measure and the theoretical gaps it can address. Its built-in weight by country also makes it a globally standardised measure with real-world applications and has demonstrated a realistic bases for poverty thresholds. This makes it better than just obtaining the income data of individuals and making an arbitrary threshold. Although Bradshaw and Holmes (2010) do suggest the use of multiple poverty variables, the others that are available and used in the MCS are more subjective and would not be as robust a measure alone as income is. Also, adding all of the poverty variables together into a composite which differ qualitatively, and have different levels of robustness, may also make the measure more at risk of issues with internal reliability. Further, it would be difficult for establishing proper interpretation of if all or only some of the composite is driving moderating effects. As such, the OECD 60% threshold measure was used alone.

Theoretical considerations

Theoretically, it has good links to language and school readiness, and current research indicates it has a direct effect on school readiness (as seen in section 3.1.2). It has also yet to be explored in a moderation analysis for the relationship between language and school readiness. Therefore, it had good potential to examined as a moderator.

Methodological considerations

This measure is used in MCS studies already as a valid measure to determine poverty (e.g. Bradshaw & Holmes, 2010). It would be expected that in the general population, 60% would fall into the above poverty median, and 40% would fall below. However, because the MCS oversampled families from poorer and more deprived areas, those who are defined as being in poverty may be higher. This was not found, with the proportions of families above the poverty threshold closer to 70% (n=6,946, 69.22% of non-missing cases in the in-scope sample. This could not be explained by the number of children missing data, as this was such a small proportion of the in-scope sample (n=66, 0.9%). When looking at the full sample at wave 2 (n=15,576), these proportions are very similar (below n=5,082, 32.63%; above n=10,307, 66.17%; missing n=187, 1.2%). As such, the MCS was unable to obtain or maintain a sample which would be 1:1 with the OECD estimated threshold. However, there is still a relatively large sample of families in poverty represented in this sample, and the sample size is large and represents over 2000 children from all over England. Therefore, although interpretation will have to come with the caveat that comparisons cannot be made with 1:1 representation, it can still represent many children in England.

Initial school readiness: Bracken School Readiness Assessment – Revised

Measure description overview, and selecting score type

The Bracken School Readiness Assessment - Revised (BSRA-R) is a standardised test created by Bracken (1998) and like the BSRA-R, completed with the children via CAPI in the MCS. It is a subset of the Bracken Basic Concept Scale-Revised which was designed to assess concept acquisition (Centre for Longitudinal Studies & GfK NOP Social Research, 2006a). It is made up of 88 items in six subtests aiming to test the knowledge and understanding of six concepts; colours, letters, numbers, sizes, comparisons and shapes. Items involved children being shown pictures and asked a question relating to one of the six concepts.

Similar to BAS-2 Naming vocabulary, the standard score was chosen for the same reasons. There was also the normative classification that could be used, which places children into a categorical grouping based on their standard score; ranging from: ‘Very delayed’, ‘Delayed’, ‘Average’, ‘Advanced’ and ‘Very advanced’. Although this can present a meaningful qualitative indication of ability, it does have a disadvantage of losing detailed information provided from the standardised score. As such, the standard score was used.

Theoretical considerations

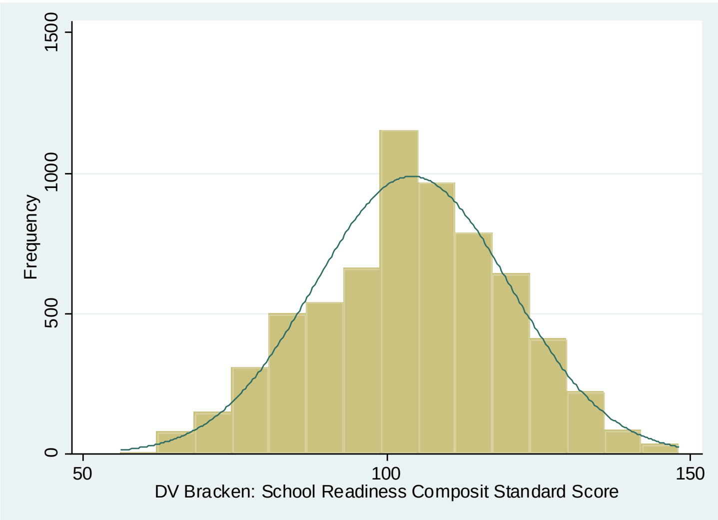
BSRA-R was chosen as a covariate to control for initial school readiness, so that the association of the predictor and moderators could be seen more clearly. Multiple prior studies using the MCS have used the BSRA-R as a measure of school readiness (Camacho et al., 2019; Cullis & Hansen, 2008; Hobcraft & Kiernan, 2010). However, it should be noted that the BSRA-R covers more basic (and only academic) abilities than the FSP, which will not cover all abilities and variance. However, it would be difficult to divorce the BSRA-R completely from being a measure of school readiness, as academic skills are still an important part of school readiness, and as previously mentioned has been linked with other educational and teacher-based assessments. In addition, it has been used as a precursor to the FSP assessment in MCS studies, and strongly predicts its outcomes, or is used to help account for school readiness at age 5 (e.g. Cullis & Hansen, 2008; Hobcraft & Kiernan, 2010).

Methodological considerations

The BARA-R demonstrates high reliability as is scored based on the same procedures and via a standardised manual each time. This means the measure is reproducible and the data produced is reliable. This measure demonstrates high representativeness as it can be used for children between the ages of 3;0 to 7;11, and is normed from over 1100 children living in the USA between two and a half and eight years. The sample included a variety of ages, assigned at birth gender, ethnicity, region and parental education. Also like BAS-2 NV, the ages are binned in three month age groups (although bins are slightly different ages). It should also be noted that the test was slightly affected in its generalisability, as it could only be completed in English and Welsh only, so children with little to no English or Welsh language skills could take part (Centre for Longitudinal Studies & GfK NOP Social Research, 2006b). In addition, scores had to be obtained from all six subscales for a total to be obtained.

Overall, 494 children did not have total BSRA-R scores in the final sample. The reasons were the test not being carried out (n=94), one or more sub-tests not being completed (n=393) and age unknown (n=7). Other than not speaking English/Welsh, it is unclear how many did not complete due to this, or what the coded reasons refer to (other than not having age data). So it is unclear who the missing children represent. However, when examining the data for the current sample, it shows that the mean(SD) score of the sample is 104.08(16.14), and scores are normally distributed (see histogram in figure 3.10) with a skew of -.14 and a kurtosis of 2.59. In the total UK and full English sample, the mean score for both was 105.6 (George et al., 2007). This means that the sample's average is around the measurement and full MCS sample average, and the full range of the measure is represented well in the current sample.

Figure 3.10. Histogram of sample BSRA-R scores (n=6,518)



Note. As this was a standard score, when considering comparisons to the normed population, the average score achieved would be 100, with +/-1SD on the score equating to +/- 15 points.

Age in months: Main responder reported (CAPI)

Measure description

Age data (in days) obtained in the MCS were calculated by the CLS based on their birthday date and the date of the interview via the CAPI questionnaire.

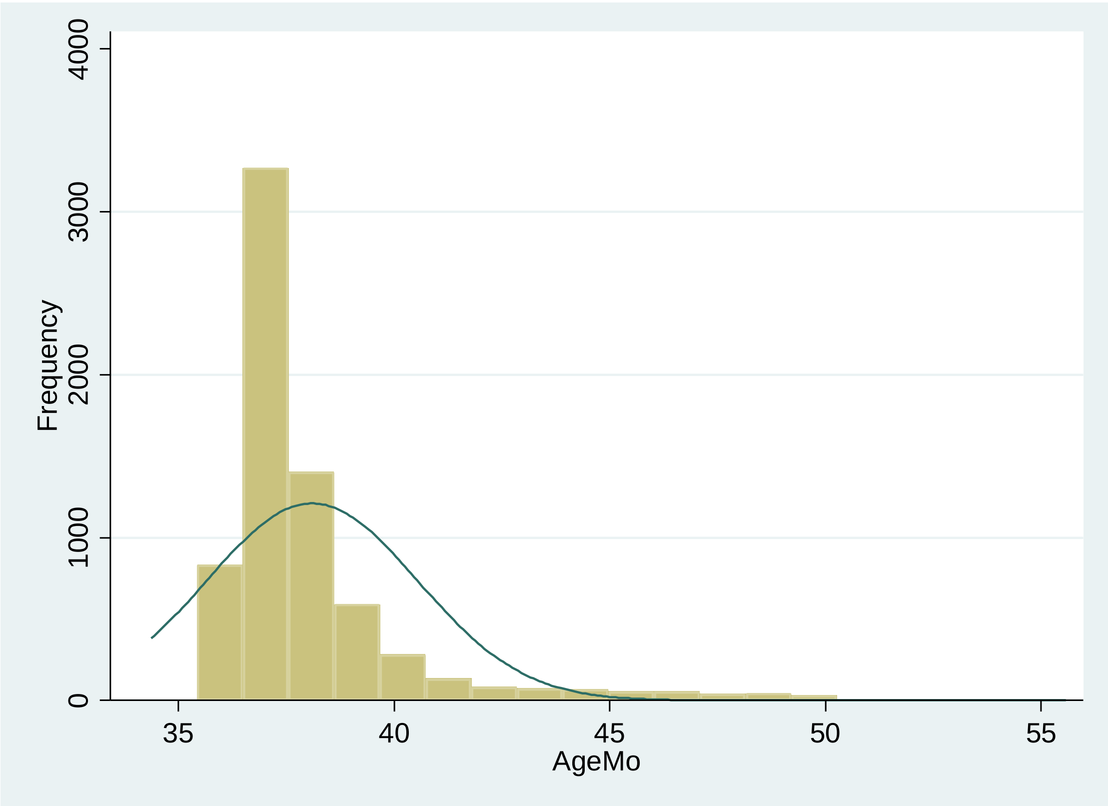
Theoretical considerations

Age is considered an inseparable to the scoring of BAS-2 Naming Vocabulary and FSP, and so produces a confounding effect. For instance, BAS-2 Naming Vocabulary T-scores are calculated in part from the age of the child (Connelley, 2013), and certain categories in the FSP may have more mastery depending on the season the child was born in (Department for education, 2020; Hobcraft & Kiernan, 2010). The Department for Education state that being older and potentially more mature allows children to be more “highly active and more likely to demonstrate what they know, understand and can do in situations that are sympathetic to this inclination” (p.16, 2020). Therefore, it would be extremely difficult to separate the ‘true effect’ of the independent variable (i.e. it is not the expressive vocabulary ability, but actually the age) and the ‘true outcome’ (i.e. the outcome is actually based on the age of the child) from the effect of age (although the variables being at different times may make this more difficult to ascertain). As such, it may be more informative to treat age as a covariate (with the aim of adjusting for the confounding effects of age). Connelley (2013) also suggests using age as a covariate, as the three-month bands that children are scored into may not necessarily be sensitive enough to pick up age differences. Therefore, a higher sensitivity than three months will be useful to address this concern.

Methodological considerations

Data was reported in days. Except for human error, acquiring and calculating this data appears to have little risk of being unreliably obtained, and the method to obtain and calculate age appears to be easily reproducible (with the correct level of dataset). Further exploration was done to understand the sample in terms of their age, and to consider the representativity, and interpretability of the variable. The sample mean(SD) age was 38.1(2.45), or just over 3;2 years, and ranged between 34.37 months to 55.57 months. The data were not normally distributed, with a skew of 2.91 and a kurtosis value of 12.64. These findings suggest that there is a part of the distribution which is extremely heavily distributed to a particular grouping of ages, with only a few cases outside of this. When examining the histogram (figure 3.11), data is mostly clustering around 36-41 months (i.e. 3;0-3;11 years old). 6,826 out of 7,012 (97.35%) of the children are around this age. However, this is expected as the data of the respondents/ children needed to be collected around specific ages as closely as possible (i.e. at 3 years old for wave 2). A small number of children aged before 3;0 (n=74) and 4;0+ (n=108) were also in the sample, but did not differ on expressive vocabulary (F(2,7005) = 2.05, p=.55) or school readiness (F(2,7005) = 0.60, p=.13) scores to those aged between 3;0 and 3;11. All but 4 children (0.06% of sample) had age (in months) noted. As such, the data reflects children who are predominantly at age 3, and so indicates that the in-scope sample represents the targeted age at this wave very well, would not be affected by missing data.

Figure 3.11. Histogram for ages in months (n=7,008)



Maternal education: Mother reported NVQ level (CAPI)

Measure description: Overview and excluding children with mothers that have overseas qualifications

Measurement of maternal education in the MCS reflected mothers' highest qualification from a list of academic (higher degree to GCSE) and vocational (professional qualifications at degree level to NVQ/SVQ/ GSVQ level 1) qualifications. There were also options for none or overseas qualifications. Data for qualifications was taken if mother had provided data via being a main or partnerresponder, or if data was onlyavailable from a fed-forward variable from wave 1. The specific variable used was derived by the CLS, which collapsed both the academic and vocational options. The NVQ levels were as follows:

1) **NVQ level 5** = higher degree;
2) **NVQ level 4** = first degree, diplomas in higher education, professional qualifications at degree level and nursing/other medical qualifications;
3) **NVQ level 3** = A/AS/S levels, NVQ/SVQ/GSVQ level 3;
4) **NVQ level 2** = O level/ GCSE grades A-C, trade apprenticeships, NVQ/SVQ/GSVQ level 2;
5) **NVQ level 1** = GCSE grades D-G, NVQ/SVQ/GSVQ level 1;
6) **Overseas qualifications only** = Other academic/vocational qualifications (including overseas);
7) **None of these qualifications** = None of these (option in both academic and vocational lists).

Children whose mothers had overseas qualifications were removed. This is because it was impossible to rank like the other options (i.e., higher level=higher qualification), and it was likely qualifications obtained overseas could have been equitable to any level of qualification. It would also not be theoretically sound to include these overseas qualifications children as missing, as these would be mothers with distinct qualifications. This was also done in other studies using MCS data (e.g. Camacho et al., 2019; Cullis et al., 2008; Hobcraft & Kiernan, 2010). So, a caveat to consider bias introduced by removing overseas qualifications will need to be considered when interpreting the findings later.

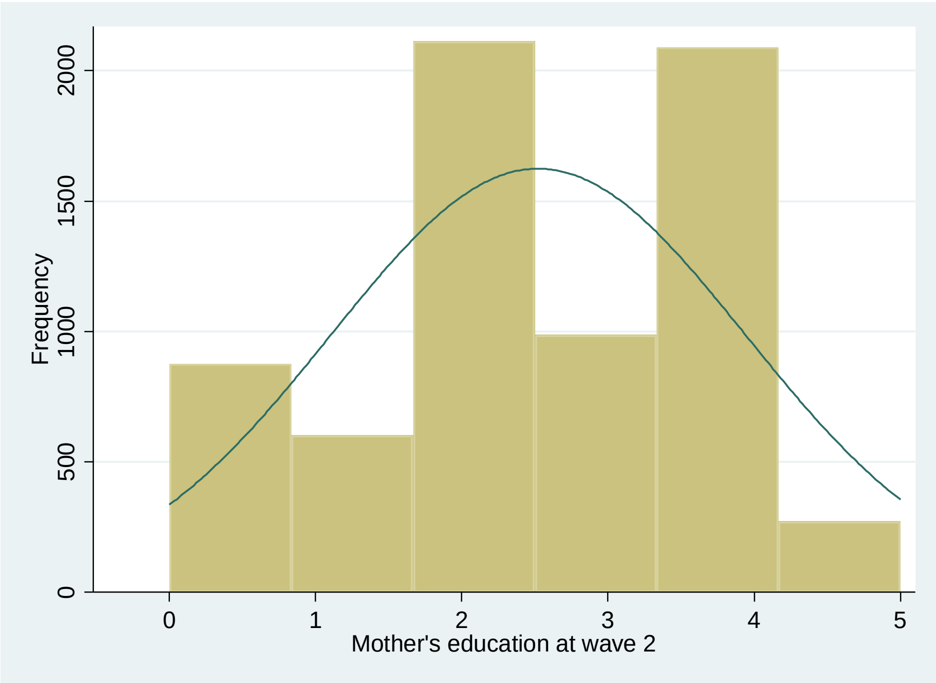
Theoretical considerations

Maternal education appears be an indirect rather than direct social characteristic. Jackson et al. (2017) notes that the reason maternal education is a strong indicator of SES and seems to link so well with children's development is because it predicts other resources that a family has, and which then in turn predicts wellbeing for children. They state that research has found lower educational levels to be associated with mental health problems, economic insecurity and a higher incidence of unstable family environments. Dickerson and Popli (2016) and Harding et al. (2015) also claim that maternal education shapes income and 'parental investment', which in turn provides a direct provision of resources children need for education. With this in mind, maternal education may mediate variables which moderate the relationship between language and school readiness. This may cloud the interpretability of its effects if not considered properly. As such, in this analysis it may serve better as a covariate in order to account for its indirect effects on other variables.

Methodological considerations

The measure was valid and representative, as it reflected all formal qualifications in the UK at the time, and had well-defined differences in levels (e.g., the separation of GCSE grades into NVQ 1 (GCSE grades of D and below or equivalent) and NVQ 2 (GCSE grades of A*-C or equivalent) relates to how the government takes note of passes at A*-C in official education league tables). This was asked in the CAPI interview, meaning data collected is easy to replicate. In the selected sample, the median level and most obtained level of qualification was NVQ level 2 (n=2,113, 30.44% of non-missing cases) followed closely by NVQ level 4 (n=2,090, 30.11% of non-missing cases). NVQ level 5 was the least obtained by mothers (n=272, 3.92%), followed by NVQ level 1 (n=603, 8.69%), No qualifications (n=876, 12.62%), and NVQ level 3 (n=987, 14.22%). Therefore, a large number of mothers were at least high-school (or equivalent) educated. This reflects similar proportions in official government statistics for educational levels around the same time (Department for Education and Skills, 2004). This indicates this variable and the current sample is representative of the general UK population, and would suggest that responders reliably reported their qualifications. Other than overseas qualifications, only 71 of the final sample had missing data, meaning proportions of qualifications were unlikely to be influenced by missing cases.

Figure 3.12. Histogram of maternal education NVQ levels



Note. Each bar represents 1 level, with the first being 'no qualifications' to the 6th being NVQ level 5 (n=6,941). As outlined, those with mothers with overseas qualifications were excluded (n=225).

Behavioural difficulties: Strengths and Difficulties Questionnaire - Main responder reported (CAPI)

Measure description: Overview and selecting total difficulties

The Strengths and Difficulties Questionnaire (SDQ) was used to measure behaviour. The following information about this measure is provided by the official SDQ website (Youth in Mind, 2016). The SDQ is a standardised screening test that consists of 25 items split into 5 subsections of 5 questions each, being emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and prosocial behaviour. Each question has a three-point response scale ('Not true': 0, 'Somewhat true': 1, and 'Certainly true': 2), meaning each subscale has a minimum of 0 and a maximum of 10. Behaviour is split into a four-band system, with total scores (excluding the prosocial scale) being ‘close to average’ if between 0-13, ‘slightly raised’ if between 14-16, ‘high’ if between 17-19 and ‘very high’ for 20+.

For the analyses, the SDQ total difficulties score was used. It was decided not to include the prosocial behaviour subscale. This was because the focus of the question was on difficulties rather than positive behavioural traits. It was also decided not to split the measure into ‘internalising problems’ (using the items from the emotional symptoms and peer relationship problems subscales) and ‘externalising problems’ (using the items from the conduct problems and hyperactivity/inattention subscales) as recommended by Goodman et al. (2010). This was not done as the its inclusion was about the level of behavioural difficulty rather than providing a clinical diagnosis. The bands were not used as total scores are better than categories for richness of data. However, thresholds were used as a guide for interpreting score values.

Theoretical considerations

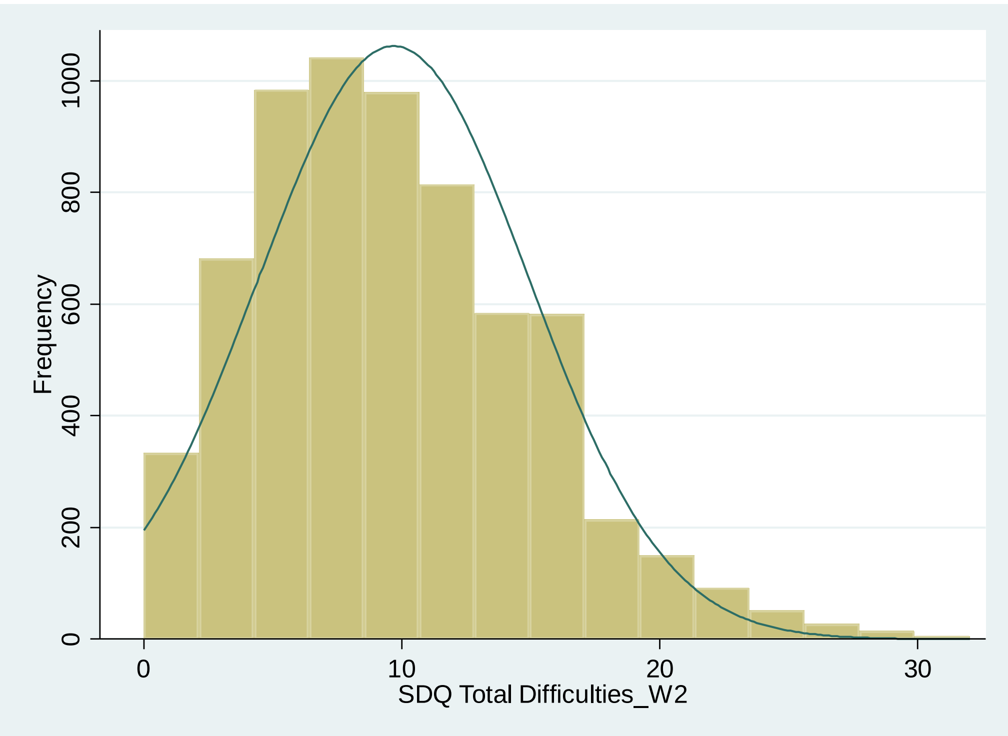
Theoretically, it has good links to language and school readiness (as seen in section 3.1.2). It has also yet to be explored in high-quality moderation analysis (as noted especially in the systematic review). Therefore, it had good potential to examined as a moderator.

Methodological considerations

It was normed using a large community sample of The British Child and Adolescent Mental Health Surveys, which was representative of the UK population (Goodman et al., 2010). SDQ data for behaviour was considered, as its standardised nature and questionnaire format allows it to be easily replicable and simple to complete. It is commonly utilised in research and different socio-cultural settings, which demonstrates wide applicability and robustness (Goodman et al., 2010). It is also supported in its validity across gender, age, alongside other disorders, and parental education by research, and is a good predictor of adolescent behavioural issues when assessed at pre-school age (Bjerke et al., 2018; Dahlberg et al., 2019; Maurice-Stam et al., 2018; Van Roy et al., 2008; Vugteveen et al., 2021).

It was decided to not have behaviour as the moderator, due to the proportion of children with behavioural issues, and the number of missing cases in the sample. Those scoring high and very high (484 and 425 respectively), make up 13% of the sample population. In contrast, only 5% of children were reported to have clinical behavioural issues in the UK in 2004 (National Institute for Health and Care Excellence, 2021). This higher proportion may be due to the SDQ being a screening measure and parent reported, which is usually then followed up by more thorough assessment. It may also indicate that the MCS sample is not representative for behavioural issues, and is a 'high risk' sample. However, 5,290 of 6,558 children were scored 'close to average' and 813 were 'slightly raised'. The average total score was 9.68(5.25), and the skewness (.75) and kurtosis (3.5) suggested a slight bias towards lower scores. Therefore, it is likely the former rather than the latter issue. As such, the variable may still give a useful indication of behaviour in the sample, but may be less accurate than other measures selected. Furthermore, there are 454 data points missing, making it one of the variables with the largest missing data in the sample, and represents nearly 7% of children in the final sample. As such, these could introduce biases which will affect the measure's implications as a moderator.

Figure 3.13. Histogram of SDQ total difficulty scores (n=6,558)



Home Learning Environment: Main responder reported (CAPI)

Measure description

The Home Learning Environment (HLE) index directly measures the learning practices at home. The HLE index was created and examined by Melhuish et al. (2001, 2008) in order to highlight specific areas of learning activities that contribute to a child's cognitive development. There has been a number of HLE measures, but the one used in the MCS was the HLE index (Melhuish et al., 2001;2008). De La Rochebrochard (2012) provides an outline of the HLE index for the MCS. It consists of seven items (being read to, going to the library, playing with numbers, painting and drawing, being taught letters, being taught numbers, and engaging in songs/poems and rhymes) with a maximum score of 7, and this is based on two questions asking if anyone does the activity with the cohort child, and if so, how often (for read to, the who question is split into if the main responder or someone else reads to the child). The ranges of time have different frequencies depending on the item, but generally range on a six- or seven-point scale from never being done to always being done (de La Rochebrochard, 2012).

Theoretical considerations

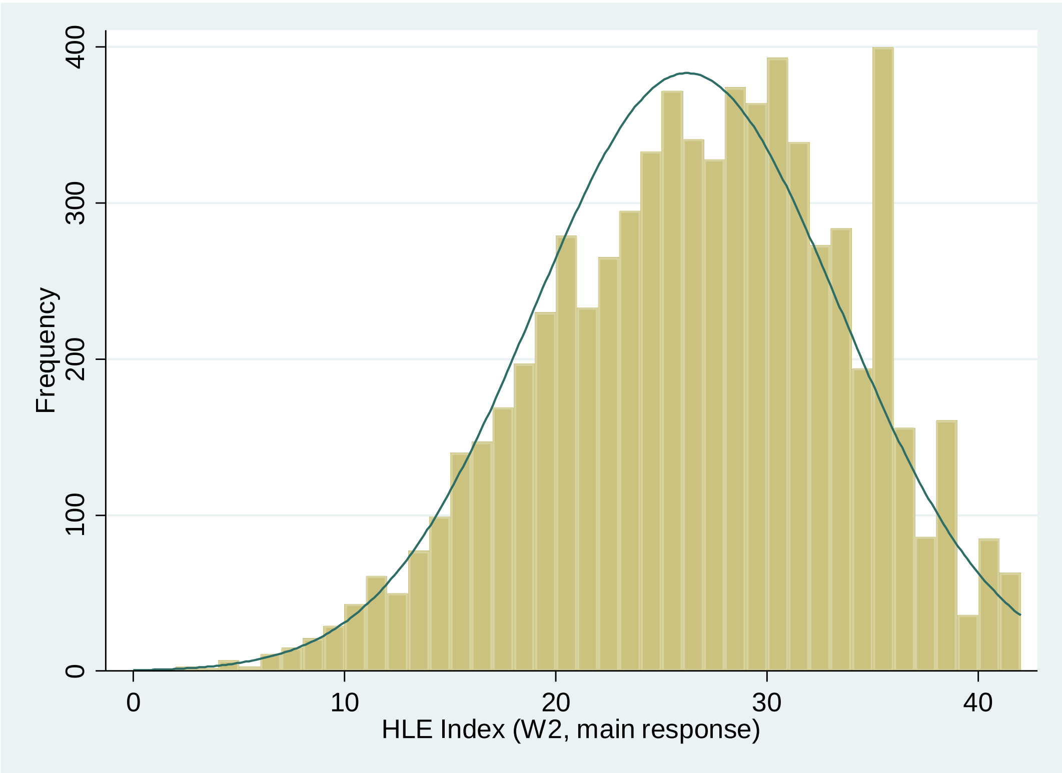
HLE was not chosen as a moderator because research suggests that it may more specifically improve language skills rather than all domains of school readiness (Rodriguez et al., 2011; Son et al., 2010). As such, like age this measure may cause differences in both language and school readiness, which would make it a confounding variable. Using the MCS Kelley et al. (2011), the HLE index did uniquely explain variance in reducing differences seen caused by income for behaviour, but not so much for BAS-2 Naming Vocabulary and school readiness (BSRA-R) at age 3. This may be explained by Dickerson & Popli (2016), who found that income influences the amount of HLE parents 'invest', and also may be due to parents changing their 'investment' based on the child's performance. This evidence suggests that the HLE may have a predictive impact on FSP, but may be influenced itself by material resources and parental views of initial child ability, as seen with the research showing it predicts initial language potentially due to its literacy building elements, but also may be tempered by parental perception of abilities, meaning it could potentially have a confounding effect on the relationship between language and school readiness, although this is not 100% clear. As such, the HLE may be better in the current study as a covariate, in an attempt to clarify if it has a unique contribution to FSP outside an economic factor and language ability.

Methodological considerations

Data for the HLE is based on the Effective Provision of Pre-school Education (EPPE) study (Melhuish et al., 2008), including 141 preschool centres (2,857 children) randomly chosen in six local authorities and considered to represent the demographic characteristics of England overall, meaning data is fairly representative to the wider population. However, it is unclear what groupings may have been included in 'norming' the measure.

In the MCS, data for all activities but playing with numbers was collected. This was suggested by Melhuish to not be an issue because the shape of the distribution rather than the range is more important, and that actually learning something at home is more important than the nature of the activity. As such, the HLE appears to be a valid measure of the amount of learning happening at home. However, this response may raise questions about what specific tasks are useful, but research from Melhuish et al. (2001, 2008) examined the seven prior mentioned activities at preschool age alongside others (play with friends at home, play with friends elsewhere, visiting relatives/friends, shopping, TV, eating meals with family, regular bedtime) for at the beginning of preschool for children's late general cognitive ability (measured by the BAS-2) at age 7 and 8. It was found that only the seven in the current index had a significant and positive impact, and the distribution of results were normal (mean = 23.42; SD = 7.71). As such, the HLE was considered to be a standardised and accurate measure due to its representative sample and findings for academically based skills (de La Rochebrochard, 2012). In the MCS, the HLE index for the whole cohort was also shown to have an approximately normal distribution for the HLE index (i.e. has a normally distributed shape, despite having some outliers at the bottom end of the total), even though the mean score was higher than Melhuish et al.'s (2008) data (mean = 25.8, SD = 7.39; de La Rochebrochard, 2012). In the selected sample, the mean(SD) was 26.22(7.25), and the histogram (skew of -0.29 and a kurtosis of 2.69) presented in figure 3.14 compared to those in de La Rochebrochard (2012) are almost identical. Therefore, the HLE index score shows a strong reliability, and in addition to the general MCS sample, the selected sample is representative of the general population.

Figure 3.14. Histogram of total HLE index scores (n=6,960)



Maternal age at birth of cohort child: Mother reported (CAPI)

Measure description

Maternal age at birth is a relatively straightforward variable, and was obtained by calculating the difference between mother's age and child's age at wave 2. In the majority of the literature, age at birth with MCS data is split between teenage age, 20s, 30s and 40+ (Camacho et al., 2019). But as with other measures, this was left as a continuous variable so that data was not lost in the analysis.

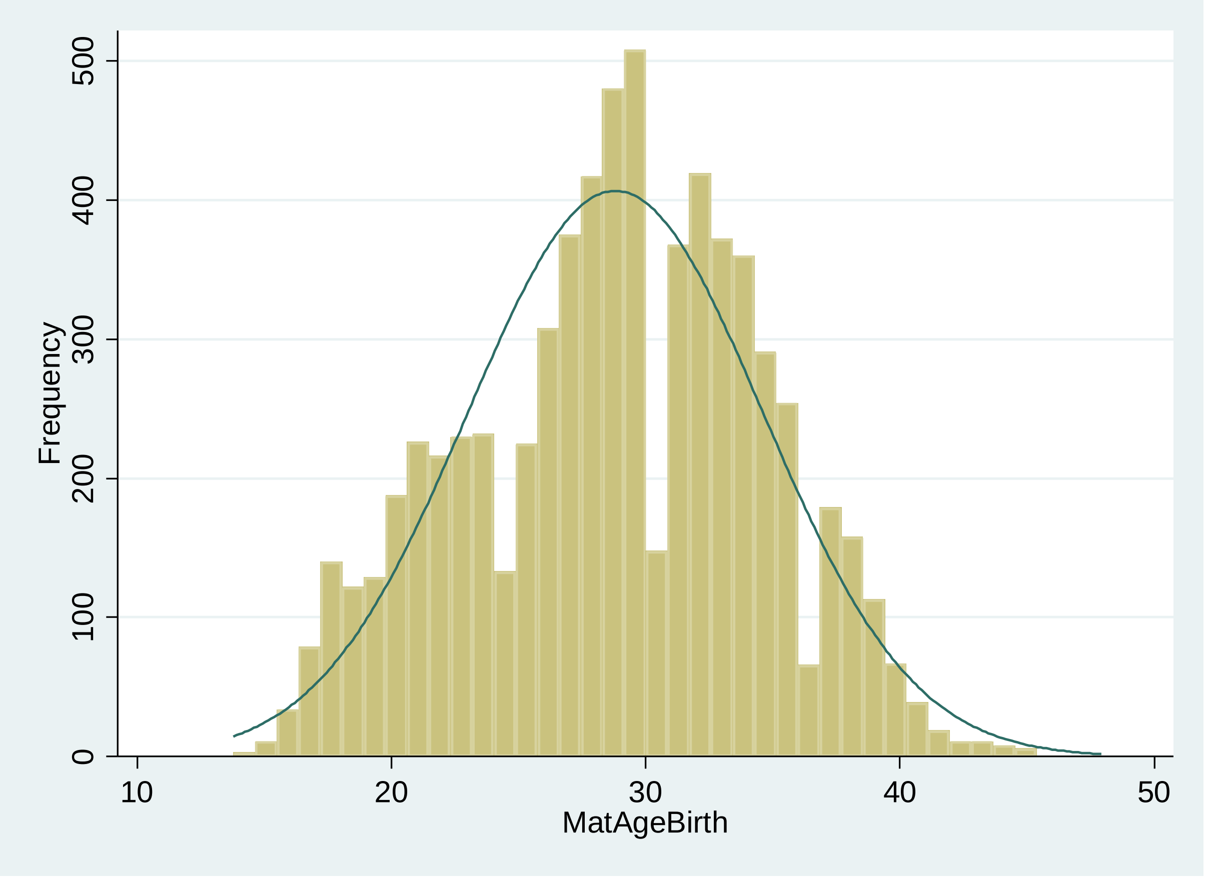
Theoretical considerations

Maternal age at birth was eventually chosen as a covariate rather than a moderator because research indicates it may be situated in a complex system of other variables like SES. For example, young mothers are likely to be from disadvantaged backgrounds (Hawkes & Joshi, 2012). However, it is unclear whether being disadvantaged produced the higher risk for being a young mother, or being a young mother produced the risk of being in disadvantage. While some research seems to state that maternal age is separate from factors such as income (Hobcraft & Kiernan, 2010), others indicate being an older mother is a common characteristic of being advantaged (Gosis et al., 2017; Schulkind & Sandler, 2019). Specifically, research suggests being younger as a mother may make you more likely to interrupt schooling, less likely to be married, and less likely to earn good income. These will in turn provide less learning resources for their child. Therefore, to try and account for this complicated relationship, maternal age at birth of cohort child is used as a covariate.

Methodological considerations

Although not strictly a standardised measure, like the child's age, there is little to question in terms of its validity or replicability (i.e. obtained via CAPI, calculation available in syntax 12.3). In the final sample, the range of ages when giving birth to cohort child were between around 13 to nearly 48 years old, with the largest group being mothers aged 20-29 (n=2,277, 48.60% of non-missing cases), followed by 30-39 (n=2,795, 40.23% of non-missing cases). This was then followed by teenage mothers (n=677, 9.74% of non-missing cases) and mothers over 40 (n=99, 1.42% of non-missing cases). The mean(SD) age when the mothers gave birth to the cohort child was 28.81(5.82) years old, and ages were normally distributed with a skew of -12 and a kurtosis of 2.46 (see figure 3.15). When viewing the average mother age for England and Wales in 2004, the average was 27.1 years (ONS, 2014), suggesting that the final sample is around 1.7 years older. However, this difference is relatively small and age does not fall into a new age bracket (i.e. 30-39 years). Together, these indicate that maternal ages collected are what could be representative of a general population. There were a very small number of missing cases (n=64), therefore, it is unlikely data will be influenced by missing cases.

Figure 3.15. Histogram of mothers' ages when cohort child was born (n=6,948)



Maternal mental health: Kessler-6 - Mother reported (CAPI)

Measure description

The measure used to assess maternal mental health was the Kessler 6 scale (K-6; Kessler, 2003). In the MCS documentation, Johnson et al. (2015) describes the measure, and further details are provided by the National Comorbidity Survey (2003). The K-6 is a 6-item short form (a 10-item form is available but not featured in the MCS) which screens for serious mental health conditions. Via CAPI, responders self-reported the frequency over the past 30 days on feeling 1) depressed, 2) hopeless, 3) restless or fidgety, 4) everything was an effort, 5) worthless, and 6) nervous. The maximum score is 24, and ranges from 0 to 4, with options being ‘none’ (0), ‘a little’ (1), ‘some’ (2), ‘most’ (3) or ‘all of the time’ (4). Once completed, scores can indicate low (0 to 3), moderate (4 to 6), high (7 to 12), or very high (13 to 24) risk of psychological distress and serious mental illness.

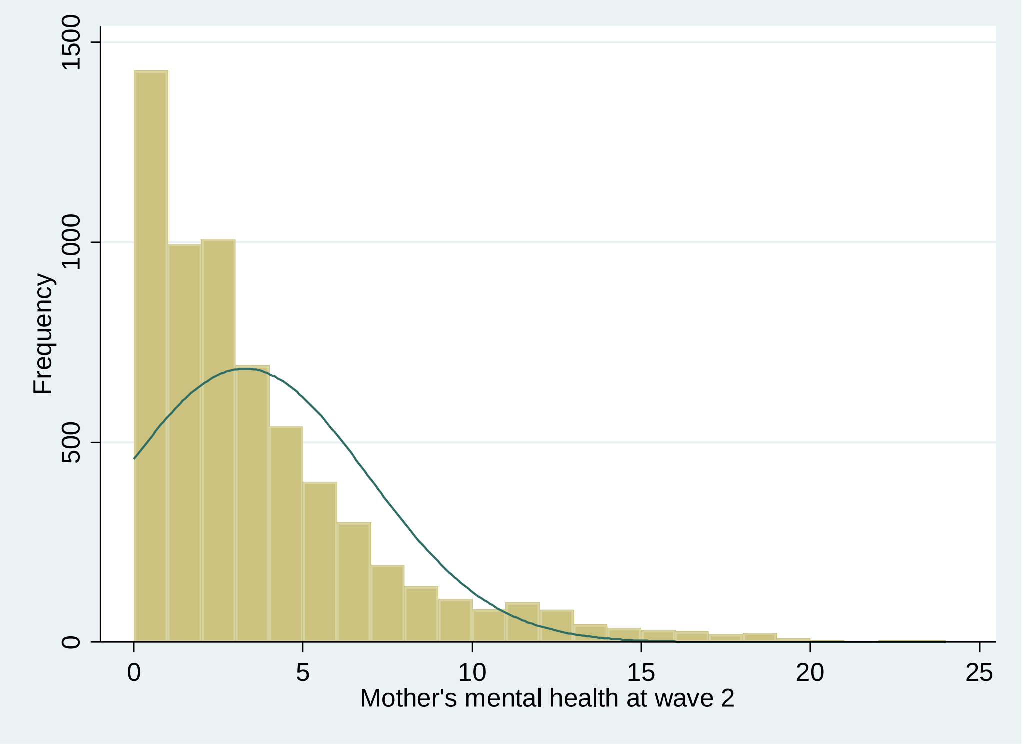
Theoretical considerations

Kiernan and Mensah (2009) found that maternal mental health and poverty both weakened the effects of one-another on school readiness at age 3. Therefore, maternal mental health could be a unique predictor for school readiness at age 5, but it also seems to be that these two factors are linked closely, with no suggestion on what may influence the other. However, Mensah & Kiernan in their 2010 study seemed to find that education mediated the effect of mental health, but it was still a significant predictor despite this. Like maternal age at birth, this seems to have a complex relationship with areas of SES, and so may be best used as a covariate due to its unclear and potentially confounding nature.

Methodological considerations

The K-6 has been found to be have good internal and construct validity, and be a consistent measure to determine psychological disorders across multiple subgroups and in English and non-English samples (Chan et al., 2014; Furukawa et al., 2003; Kawakami et al., 2020). In the current sample, the mean(SD) score was 3.28(3.66), with a range of 0-24, and a skew and kurtosis (1.82 and 6.91 respectively) which suggested that the data was heavily weighted to low scores (See figure 3.16). When examining the numbers for each threshold, 4,123 (65.79% of non-missing cases) were low, 1,240 (19.79% of non-missing cases) were moderate, and only 904 were at or above the high threshold (11.20% of non-missing cases, 702 high and 202 very high). In a 2014 study examining UK prevalence of common mental disorders (defined as being neurotic disorders which cause marked emotional distress and interfere with daily function), estimates for mental illnesses were between 14.7% and 16.7% (NHS, 2014). However, these statistics were acquired around 10 years after this wave, included multiple demographic groups, and used a different measure (Clinical Interview Schedule –Revised) which covered more issues such as panic, compulsions and obsessions. As such, the K-6 may reflect numbers relatively similar to the UK population. However, there is also a relatively large amount of missing data (n=745), and so could influence the scores seen in the current sample. The Together, the findings for the validity, reliability and representativeness of the measure felt like a suitable measure of maternal mental health, but data may be affected by the large number of missing cases.

Figure 3.16. Histogram of Kessler-6 scores (n=6,267)



Area deprivation: Index of Multiple Deprivation - Living Environment (family data)

Measure description

The measure Index of Multiple Deprivation - Living Environment (IMDLE) is described by the Office for the Deputy Prime Minister (ODPM, 2004) to be created from using a combination of the 2001 Census and the 2001 Mid-Year Estimates, and shows deprivation across England and at 'the small area level'. The Index of Multiple Deprivation is made up of seven distinct discrete domains, with living environment being one of them. IMDLE includes two sub-domains within it that examine both 'indoors' and 'outdoors' living environments. The indoors subdomain indicators social and private housing in poor condition, and houses without central heating. The outdoors subdomain examines air quality, and road traffic accidents involving injury to pedestrians and cyclists. In the MCS, the IMDLE was split into deciles, ranging from most to least deprived. This is indicated by if it falls within the most deprived percentage of small areas (i.e. the most deprived decile shows families that fall into the top 9% of deprived small areas within England).

Theoretical considerations

As seen in the background section, area deprivation related to language development and school readiness (findings are both separate and as part of the Index of Multiple Deprivation measure), and so showed potential as an influential factor for language outcomes in the systematic review. However, as poverty was more established in the literature, IMDLE was chosen to be a covariate.

Methodological considerations

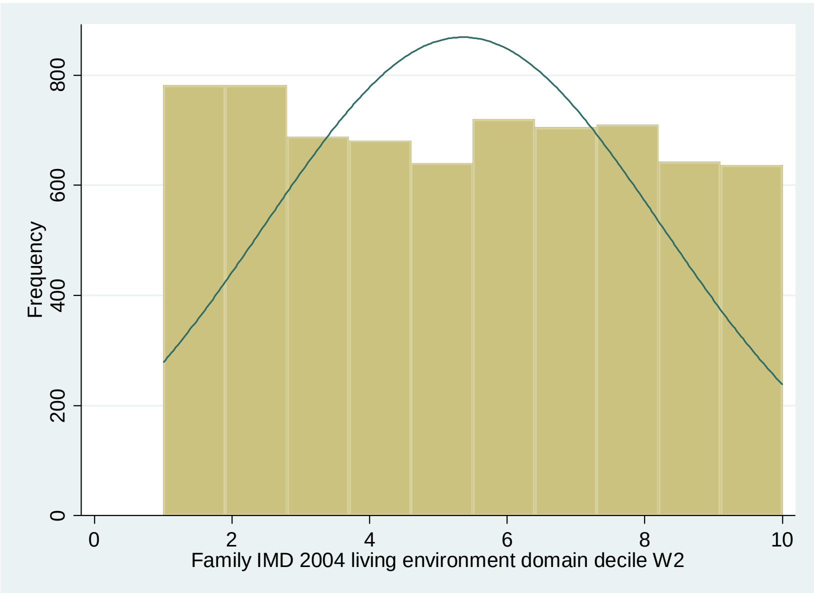
The IMDLE differentiates deprivation well, as domains can be also measured separately from the others. The IMD is a standardized UK government tool for measuring deprivation, and areas examined reflected 354 districts and 32,482 Super Output Areas (the smallest possible geographic area noted on the census) in England (ODPM, 2004), meaning that the IMD was a highly representative and representative measure for individuals living in England.

As seen in the data (table 3.13 and figure 3.17) for the final sample, deciles represented between 9 and 11%, with the biggest difference in n between deciles is 145 responses, meaning that each of the deciles were generally well represented and therefore could be applicable more generally.

Table 3.13. Proportions IMD living environment for final sample (n=6,984)

Decile	n	% non-missing cases
Most deprived	781	11.18%
10-20	781	11.18%
20-30	688	9.85%
30-40	681	9.75%
40-50	640	9.16%
50-60	720	10.31%
60-70	705	10.09%
70-80	710	10.17%
80-90	642	9.19%
Least deprived	636	9.11%

Figure 3.17. Histogram representing each decile of IMDLE (n=6,984)



Note. Each bar represents a decile, with the first bar representing the most deprived decile, and the tenth representing the least deprived decile.

As a government measure of deprivation, and with the data presented above, it is clear that it has a high level of representation to the general population. Its replicability and validity also appear to be robust as it is a government produced measure, which means data collected will be based on standardised criteria and thresholds which are accepted to measure deprivation well. Furthermore, when looking at the specific items of IMDLE, it defines area deprivation by the condition of housing, availability of central heating, air quality and road traffic accidents. This makes is a highly valid indicator for area SES because it takes into account both house and neighbourhood specific attributes where poor quality (air, housing), high incidence (traffic) or lack of (heating) are all well-established indicators of lower SES (Cakmak et al., 2016; Duncan et al., 2015; Evans & Kantrowitz, 2002; Hajat et al., 2015). It also does not just focus on one indicator, providing a more holistic picture of an area's deprivation. However, when this variable was included in analysis, no findings could be obtained for the analyses when it was MCS weighted. This may be due to the weights having an emphasis on wards, and so area data may already be accounted for. Therefore, it was excluded from the current study.

Long-term health condition status: Main responder reported (CAPI)

Measure description

For long-term conditions reported in the MCS at age 3, main responders were asked if their child had a long-term health condition, and if so, if this limited them "at play or from joining in any other activity normal for a child his/her age?". The specific conditions were also reported in the MCS, and then assigned ICD-10 criteria, but this information was not accessible without special permissions as it was restricted to secure access. Therefore, only whether the child had a condition (yes/no) and if it limited everyday activities (yes/no) could be obtained from the data.

Theoretical considerations

As outlined in the background section (3.1.2), very little research conducted, but suggestion that having a long-term condition impacts school readiness. It was also chosen based on the potential that medical issues could impact school attendance and play (which contributes to developing school readiness skills).

Methodological considerations

Due to not having access to specific diagnosis data, there were issues with using this variable. First of all, just choosing only if a child has or does not have a long-term condition has the issue that some long-term health conditions are more likely to have no severe or detrimental effects on joining in on ‘normal’, everyday activities compared to others (e.g., the everyday impact asthma versus downs syndrome would have). So one idea was to create a derived variable by combining the answers of the two questions stated previously, and then split by 1) no long-term health condition, 2) a long-term health condition that does not limit everyday play or joining in ‘normal’ activities, and 3) a long-term health condition that does limit everyday play or joining in ‘normal’ activities. In this way, it could be answered whether the severity of the group’s condition demonstrates a different relationship between predictor and outcome.

Overall, 1,099 (15.79% of non-missing cases) of the final sample reported cohort children having long-term health conditions, and of these, 198 had conditions which limited their everyday play and joining in ‘normal’ activities. This seems in line with UK estimates, although these covered all young people rather than just children at age 3. Estimates around the time for chronic physical, developmental, behavioural or emotional condition was around 15-20% (Miller et al., 2004), with more recent estimates in the UK demonstrating about 15% (1.7 million) children and young people (National Institute for Health and Care Excellence, 2019).

However, this variable was not chosen as a moderator or covariate because one insurmountable issue was how to interpret findings. Specifically, even if long-term conditions significantly moderated the relationship, some conditions which are severe (i.e. defined as limits everyday activities here) may have more influence on school readiness than others (e.g. learning difficulties versus diabetes), and there may be more of one type of condition in the variable which could influence the effect. Furthermore, it would be very difficult to parse the qualitative differences in such a varied subgroup of children. For example, while one child’s school readiness being affected could be because they are in hospital a lot (and so are missing educational opportunities), another could score poorly because they have complex educational, behavioural or emotional needs; which impacts their ability to engage in educational activities. An argument could be that separating children with long-term health conditions between those who do and do not have ability to engage in joining in ‘normal’ everyday activities will likely have some shared experiences. However, their experiences will still be very varied based on their condition, and it would be unsuitable to attempt to group children together in this way based on an unfounded assumption. As such, this variable was excluded from analysis, but long-term health conditions may be important to consider in future studies and if research teams have access to the sensitive MCS datasets.

Childcare type: Main responder reported (CAPI)

Measure description: Choosing childcare type

As the childcare data was so complex to utilise, an additional study would need to be completed to feasibly examine, or focus on all of the childcare aspects provided in the MCS. Therefore, one aspect of childcare data was selected to attempt to feasibility analyse within the scope of the current study in mind, whilst also being meaningful in providing substantial, relevant and novel knowledge. The questions in the CAPI questionnaire focused on the type, hours per week, length the provision was used, and cost.

Type was seen as the better aspect over cost and dosage for a number of reasons. Regarding cost, in the UK, policy for funding, reorganisation and offers for childcare provision to more 3-year-olds than ever in places such as nurseries, playgroups and childminders in 2004. This expansion and more accessible provision coincided with wave 2 of the MCS, meaning many of the MCS families were likely receiving the benefits of these new policies for the cohort children. As such, in the context of the MCS, it would be difficult to give any meaningful interpretation to cost when a large amount of children could be receiving similar free childcare provisions. Furthermore, when cost is a bigger issue like in the US and Australia, the level of regulation and access as opposed to the cost of childcare appears to have more of an impact on, or receipt of, quality (Cloney et al. 2016; Gorry & Thomas, 2017). As for dosage (i.e hours per week or length of time childcare used), there has been little focus on this in childcare (Zaslow et al., 2016). However, the available evidence is mixed. For example, the National Institute of Child Health and Human Development (NICHD, 2006) suggests children receiving higher non-maternal childcare hours may have more behavioural issues at kindergarten; but both the NICHD and Zaslow et al. (2016) did find that more time spent in childcare was better for cognitive and educational outcomes. However, these benefits appeared to be typically evident in certain settings associated with high-quality (i.e. formal). As such, dosage may only have an impact based on the type of childcare, and so the effect of this aspect may not be as strong of an indicator as type is.

Attempt at re-coding childcare type into derived variable

In order to attempt to make a useable version of childcare for moderation, a derived variable was created. This was because there were no single or separatable variables which would provide meaningful or accurate representations of childcare. Data for any aspect of childcare therefore needed to be drawn from several overlapping questions/data points on the CAPI questionnaire. The data was made up of 1) 'fed forward' data from wave 1 as the base information, 2) potentially multiple different childcare arrangements/types per child if reported at both waves, corrections (if incorrectly reported/noted at wave 1), 3) new data obtained via new childcare arrangements for existing families, and 4) new families included at wave 2. Something else to note was that the feed forward data had a slightly different lists and coding of childcare arrangements to those listed at wave 2, and so the arrangements had to be separately turned into the three types, and then merged together. See table 3.14 for differences between fed forward childcare arrangements and those in wave 2, and STATA syntax to construct this derived variable.

Childcare type derived variable description

Childcare was a ranked categorical variable based on three types of childcare that could have been received at wave 2. These were 'formal', 'informal local', and 'informal non-local but non-audited'. In line with literature definitions and how childcare types are audited in the UK (e.g., NICHHD, 2006); formal referred to local-authority or audited arrangements (i.e. local authority day nursery creche, nursery school, nursery or reception class in a primary or infants school, ‘special day school or nursery’ or unit for children with special educational needs); informal non-local had arrangements outside of the home which were not audited by the government (e.g. workplace/college nursery/creche, private / independent day nursery/creche, childminders); and informal local had arrangements at the family home or by relatives (e.g. grandparents, friends). ‘No care’ here was counted as informal local, as it is implied that the main caregiver would be providing the childcare. Categories like ‘other’ and ‘unspecified’ were also added to informal local as it was very unlikely it would have been any other type of formal childcare, but it was unclear if this would have been likely to be care outside of the household.

Table 3.14. Childcare type categorisation for derived variable

Childcare fed forward from wave 1		Childcare wave 2	
Type	**Arrangement**	**Type**	**Arrangement**
Informal - local and other	not working - no care 2	Informal - local and other	Looking after the child yourself while you were working at home or at your workplace
Informal - local and other	Respondent his herself	Informal - local and other	Resident husband/wife/partner
Informal - local and other	Husband Wife Partner	Informal - local and other	Grandparent in my home
Informal - local and other	Your Mother	Informal - local and other	Other relative (including non-resident parent) in my home
Informal - local and other	Your partner s mother	Informal - local and other	Care in grandparents home
Informal - local and other	Other relative	Informal - local and other	Care in other relatives home (including non-resident parent)
Informal - local and other	Friend neighbour	Informal - local and other	Non-relative (including nannies and au pairs) in my home
Informal - non-local, non-audited	Registered childminder	Informal - local and other	Non-relative elsewhere (e.g. friend, neighbour)
Informal - non-local, non-audited	Unregistered childminder	Informal - non-local, non-audited	Childminder
Informal - non-local, non-audited	Workplace College day nursery creche	Informal - non-local, non-audited	Workplace/college nursery/creche
Formal - audited arrangement	Local authority day nursery creche	Informal - non-local, non-audited	Private / independent day nursery/creche

Informal - non-local, non-audited	Private day nursery creche
Informal - local and other	not working - main care nk
Informal - local and other	not working - no care
Informal - local and other	unspecified

Formal - audited arrangement	Local Authority nursery
Formal - audited arrangement	Nursery school
Formal - audited arrangement	Nursery or Reception class in a primary or infants school
Formal - audited arrangement	Special day school or nursery or unit for children with SEN
Informal - non-local, non-audited	Playgroup
Informal - non-local, non-audited	Combined centre/Family Centre
Informal - local and other	Other

Theoretical considerations

Theoretically, cognitive and educational outcomes (including language) and school readiness (as seen in the background section). It has also yet to be explored in high-quality moderation analysis. Therefore, it had good potential to examined as a moderator.

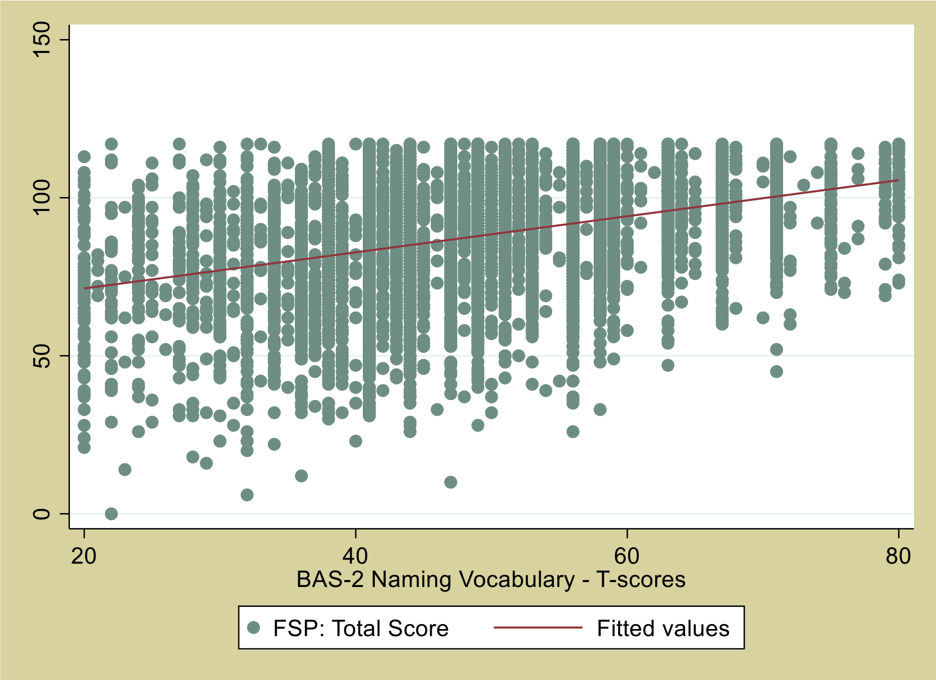
Methodological considerations

The final numbers for each category were 3,862 (57.9% of non-missing cases) for only informal – at home, 1,828 (27.41% of non-missing cases) for some informal – outside home, and 980 (14.69% of non-missing cases) for some formal childcare. Despite the work to formulate this data the variable does not have high validity. The categories made place many different types (and vaguely described types) of childcare together, and so it could not be determined if specific childcare types would be more or less influential. Therefore, placing them all together like this will not allow for a clear determination of what is influencing what. The variable was also created on assumptions in the data which could not be verified (i.e., there were some guesses to what each variable for childcare in the dataset actually were). As such, this variable was excluded from analysis due to these validity issues.

3.5.3. Appendix F. Regression assumptions graphics

Linear regression linearity and homoscedasticity (via scatterplot with fitted values)

Figure 3.18. Scatterplot for BAS-2 NV and FSP total (N=7,012)



Linear regression normality of residuals (via kernel density, Q-Q and P-P plots)

Figure 3.19. Kernel density plot for BAS-2 NV and FSP total (N=7,012)

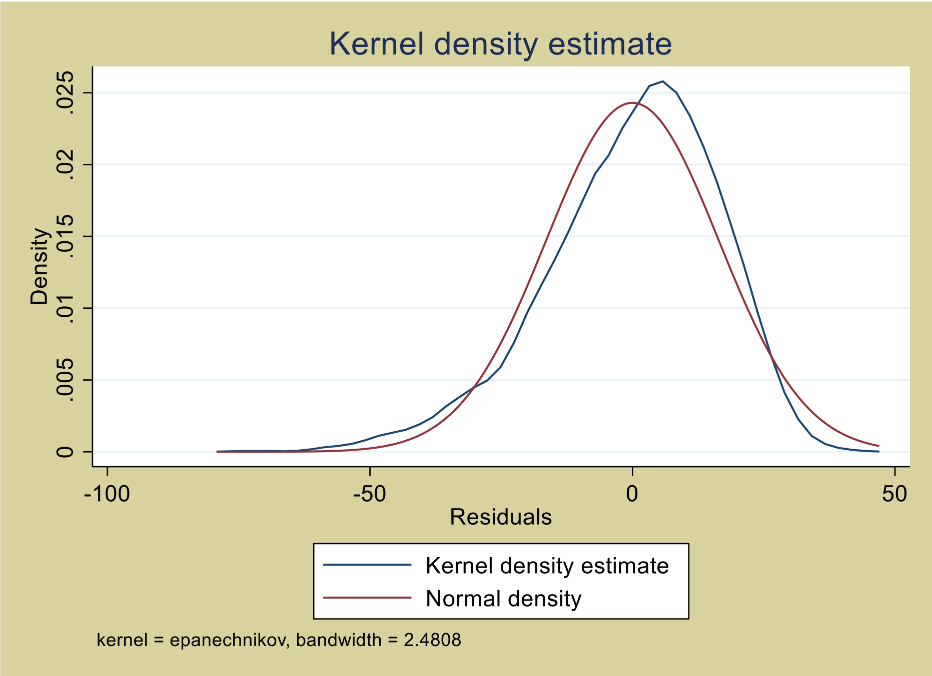


Figure 3.20. Q-Q plot for BAS-2 NV and FSP total (N=7,012)

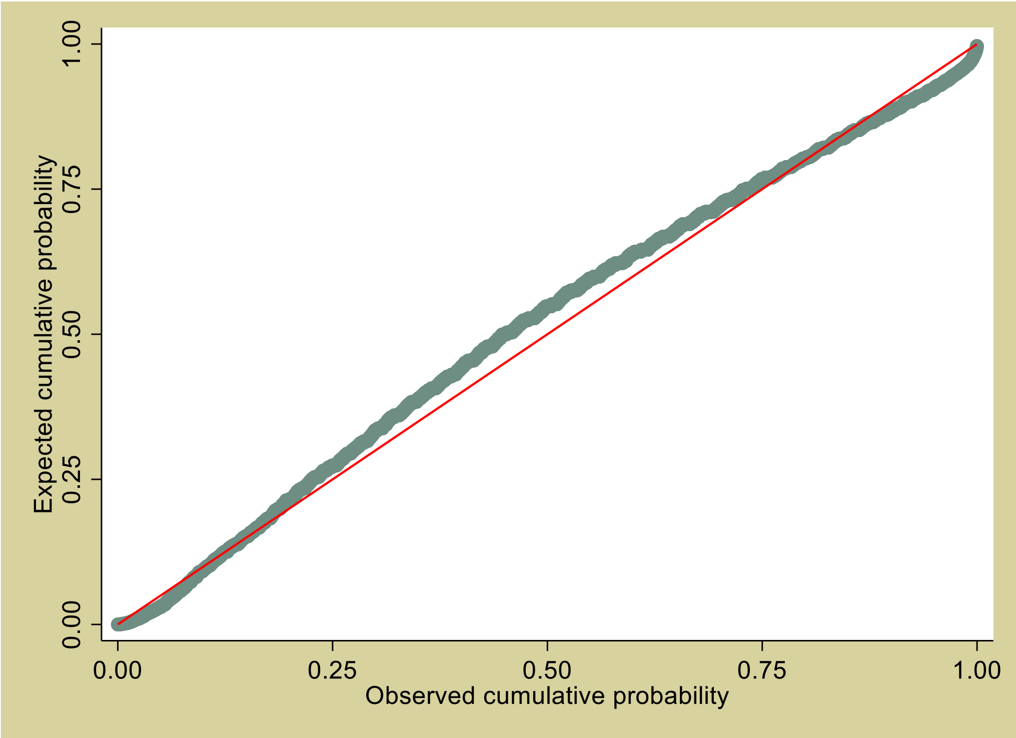
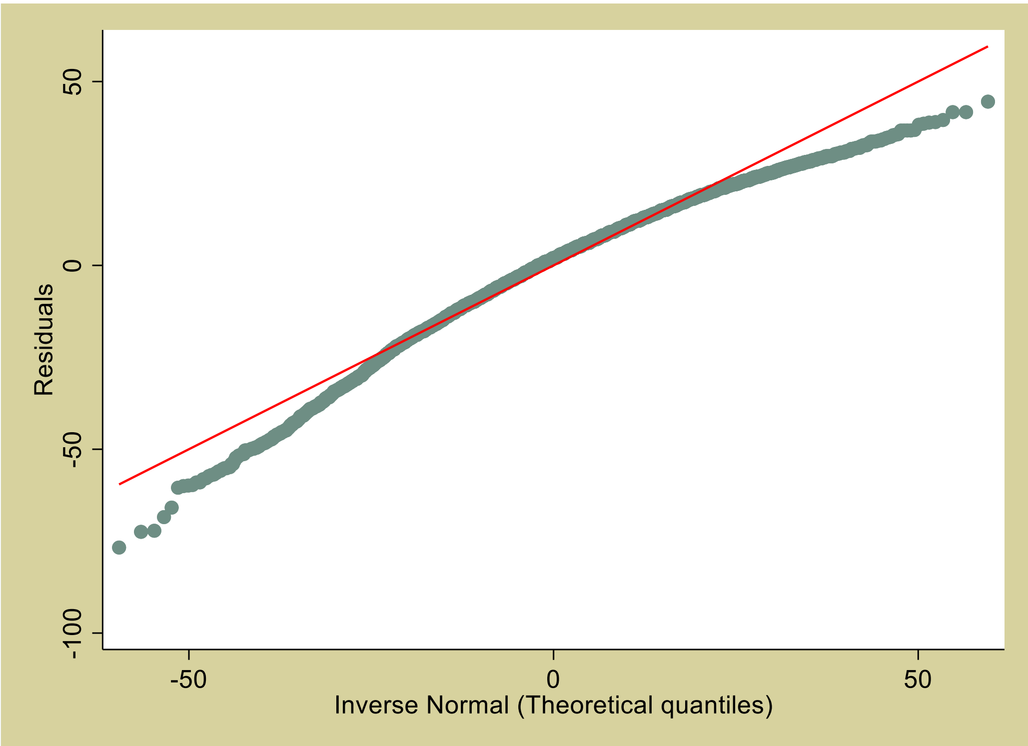


Figure 3.21. P-P plot for BAS-2 NV and FSP total (N=7,012)



Linear and logistic regression multicollinearity (via correlations and VIF)

Table 3.15. Correlation table for linear regression multicollinearity test

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1.FSP GLD	**0.73**																
2.PSE	**0.87**	**0.63**															
3.CLL	**0.95**	**0.77**	**0.75**														
4.MD	**0.91**	**0.62**	**0.70**	**0.85**													
5. KUW	**0.83**	**0.53**	**0.71**	**0.72**	**0.71**												
6.PD	**0.77**	**0.48**	**0.69**	**0.66**	**0.63**	**0.67**											
7.CD	**0.80**	**0.52**	**0.69**	**0.70**	**0.65**	**0.73**	**0.69**										
8.BAS-2 NV	**0.35**	**0.28**	**0.27**	**0.36**	**0.33**	**0.27**	**0.23**	**0.26**									
9. GAAB	**0.14**	**0.14**	**0.16**	**0.15**	**0.05**®	0.01	**0.14**	**0.21**	**0.11**								
10.Pov	**0.26**	**0.21**	**0.21**	**0.26**	**0.23**	**0.22**	**0.17**	**0.19**	**0.26**	-0.02							
11. Age	**-0.06**	**-0.05**®	**-0.06**®	**-0.06**	**-0.05**®	-0.05	**-0.06**	-0.03	0.0011	0.00	**-0.08**						
12. BSRA-R	**0.45**	**0.36**	**0.33**	**0.48**	**0.45**	**0.33**	**0.27**	**0.30**	**0.57**	**0.10**	**0.31**	0.02					
13. SDQ	**-0.25**	**-0.20**	**-0.24**	**-0.25**	**-0.23**	**-0.19**	**-0.18**	**-0.19**	**-0.25**	**-0.08**	**-0.23**	0.01	**-0.29**				
14.ME	**0.29**	**0.22**	**0.23**	**0.29**	**0.27**	**0.24**	**0.18**	**0.22**	**0.25**	0.00	**0.38**	-0.04	**0.32**	**-0.27**			
15. MAB	**0.15**	**0.13**	**0.13**	**0.16**	**0.14**	**0.12**	**0.09**	**0.10**	**0.15**	0.00	**0.32**	**-0.13**	**0.19**	**-0.23**	**0.28**		

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
16. MMH	**-0.12**	**-0.10**	**-0.10**	**-0.12**	**-0.11**	**-0.08**	**-0.07**	**-0.07**	**-0.10**	0.01	**-0.18**	0.03	**-0.14**	**0.34**	**-0.11**	**-0.10**	
17. HLE Index	**0.17**	**0.15**	**0.14**	**0.19**	**0.14**	**0.12**	**0.10**	**0.13**	**0.17**	**0.14**	**0.06**	-0.02	**0.22**	**-0.15**	**0.12**	0.01	-0.05

Note. Everything in bold is significant to the p<.001 level, except for those marked with ® which are significant to p<.01 level.

Note. These Pearson's correlations are for complete cases (n=5,718), and so would reflect multicollinearity for final models. A pairwise correlation analysis (n=7,012) showed similar results. Spearman's correlations were also completed for GLD (recommended for logistic regressions), with similar results.

Note. Correlations over .6 highlighted in orange, correlations over .4 highlighted in yellow.

Note. FSP = Foundation Stage Profile; GLD = Good level of development; PSE = Personal, social and emotional development; CLL = Communication, language and literacy; MD = Mathematical development; KUW = Knowledge and understanding of the world; PD = Physical development; CD = Creative development; BAS-2 NV = British Abilities Scale (2nd edition) naming vocabulary subtest; GAAB = gender assigned at birth; Pov = poverty; BSRA-R = Bracken School readiness assessment-revised; SDQ = Strengths and difficulties questionnaire; ME = maternal education; MAB = maternal age at child's birth; MMH = maternal mental health; HLE = home learning environment.

Table 3.16. VIF table for linear regression multicollinearity test

Variable	VIF	1/VIF
Expressive vocabulary	1.52	0.66
Gender assigned at birth (comparison = male)	1.04	0.96
Relative poverty (comparison = below 60% median)	1.34	0.75
Age	1.02	0.98
Initial school readiness	1.65	0.61
Behaviour	1.30	0.77
Maternal education (comparison = no qualifications)		
NVQ 1	1.85	0.54

Variable	VIF	1/VIF
NVQ 2	3.42	0.29
NVQ 3	2.56	0.39
NVQ 4	3.86	0.26
NVQ 5	1.57	0.64
Maternal age at child's birth	1.22	0.82
Maternal mental health	1.15	0.87
HLE Index	1.08	0.92
Mean VIF	**1.76**	**0.57**

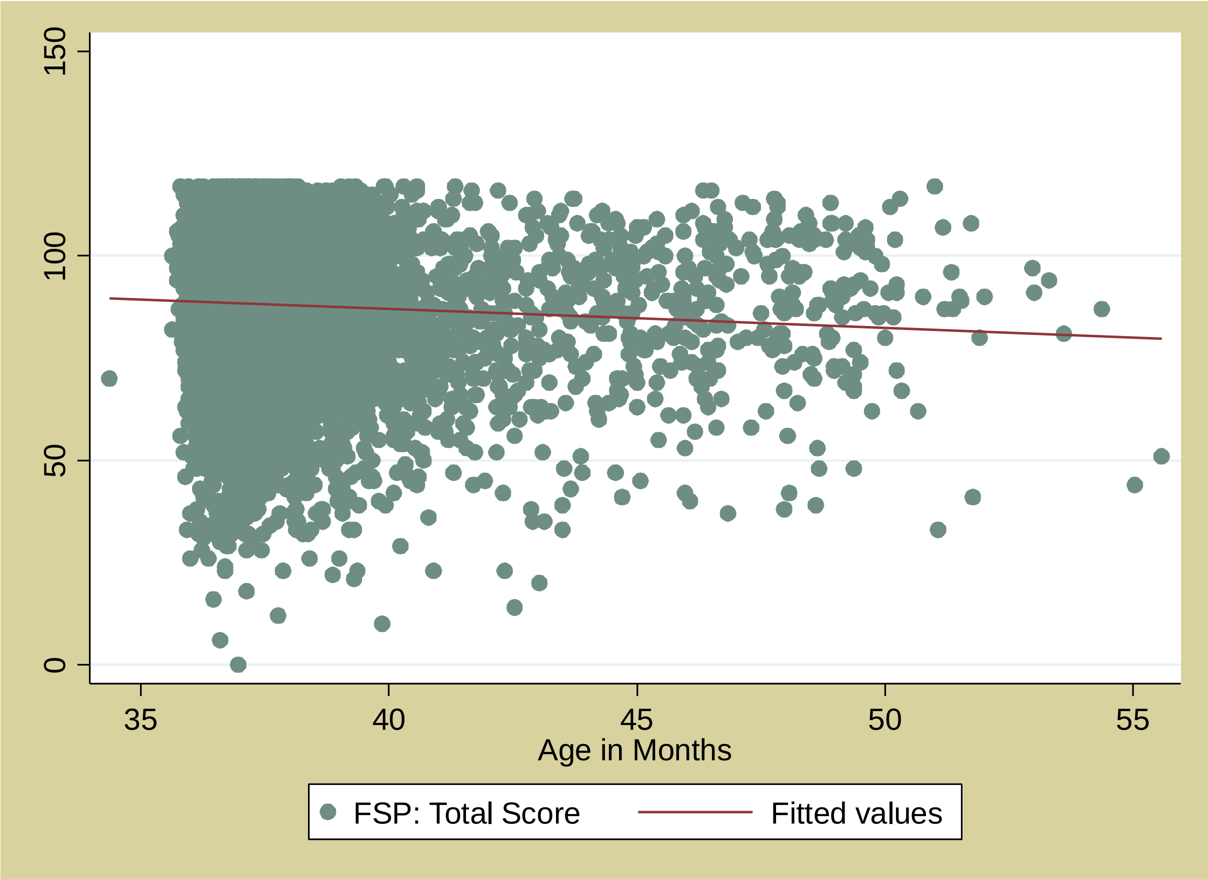
Logistic regression linearity via Box-Tidwell test

Table 3.17. Box-Tidwell test significance values for logistic regression linearity test

Variable	Model 4 significance values	Model 5 significance values
Expressive vocabulary (BAS-2 NV)	.10	.12
Age in months	.008	.008
Initial school readiness (BSRA-R)	.48	.47
Behavioural difficulties (SDQ)	.44	.44
Maternal age at birth	.08	.08
Maternal mental health	.06	.06
Home learning environment (HLE Index)	.71	.69

Note. Only continuous variables are presented here as they are the only type needed to be tested for linearity in Box-Tidwell testing.

Figure 3.22. Scatterplot showing linear relationship between FSP total and age



Note. As GLD is binary, FSP total was used as a visual diagnostic to determine the spread of data and potential influential outliers. There are some more spread out data points, but the clustering of the data for the variable seems to be more important in influencing the non-linear relationship.

3.5.4. Appendix G. Foundation Stage Profile scales moderation analyses

Profile Personal, Social and Emotional Development

Table 3.18. Linear regression models for Foundation Stage Profile Personal, Social and Emotional Development - gender assigned at birth

	Model 1						**Model 2**						**Model 4**					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0794			**R^2 = 0.0851**			**R^2 = 0.0983**			**R^2 = 0.1026**			**R^2 = 0.1729**			**R^2 = 0.1787**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.10[a]	0.00	0.09 to 0.11	0.11[a]	0.00	0.10 to 0.12	0.11[a]	0.01	0.10 to 0.13	0.12[a]	0.01	0.10 to 0.13	0.04[a]	0.01	0.03 to 0.06	0.05[a]	0.01	0.03 to 0.06
Gender assigned at birth (female)	-	-	-	-	-	-	2.65[a]	0.44	1.79 to 3.52	2.49[a]	0.49	1.53 to 3.45	2.66[a]	0.50	1.68 to 3.64	2.50[a]	0.54	1.45 to 3.55

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						Expressive vocabulary x gender assigned at birth						Expressive vocabulary x gender assigned at birth with covariates					
	N = 7,012						N = 7,012						N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	$R^2 = 0.0794$			$R^2 = 0.0851$			$R^2 = 0.0983$			$R^2 = 0.1026$			$R^2 = 0.1729$			$R^2 = 0.1787$		
Measure	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI
Interaction	-	-	-	-	-	-	-0.03[a]	0.01	-0.05 to -0.02	-0.03[a]	0.01	-0.05 to -0.01	-0.03[a]	0.01	-0.05 to -0.02	-0.03[a]	0.01	-0.05 to -0.01
Poverty (above threshold)	-	-	-	-	-	-	-	-	-	-	-	-	0.63[a]	0.13	0.37 to 0.89	0.68[a]	0.14	0.40 to 0.96
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.09[a]	0.02	-0.13 to -0.05	-0.09[a]	0.02	-0.14 to -0.05
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.05[a]	0.00	0.04 to 0.06	0.05[a]	0.00	0.04 to 0.06

	Model 1 Expressive vocabulary only N = 7,012						Model 2 Expressive vocabulary x gender assigned at birth N = 7,012						Model 4 Expressive vocabulary x gender assigned at birth with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0794			**R^2 = 0.0851**			**R^2 = 0.0983**			**R^2 = 0.1026**			**R^2 = 0.1729**			**R^2 = 0.1787**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.08[a]	0.01	-0.10 to -0.06	-0.08[c]	0.01	-0.11 to -0.06
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	0.12	0.27	-0.41 to 0.66	0.17	0.29	-0.40 to 0.73
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	0.47[c]	0.22	0.04 to 0.89	0.52[a]	0.23	0.07 to 0.97
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	0.75[a]	0.23	0.29 to 1.21	0.76[a]	0.25	0.27 to 1.24

	Model 1						**Model 2**						**Model 4**					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0794			**R^2 = 0.0851**			**R^2 = 0.0983**			**R^2 = 0.1026**			**R^2 = 0.1729**			**R^2 = 0.1787**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	1.07[a]	0.22	0.63 to 1.51	1.08[a]	0.24	0.61 to 1.54
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	1.14[a]	0.31	0.53 to 1.75	0.97[a]	0.33	0.32 to 1.61
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.01	-0.02 to 0.02	0.00	0.01	-0.02 to 0.02

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0794			**R^2 = 0.0851**			**R^2 = 0.0983**			**R^2 = 0.1026**			**R^2 = 0.1729**			**R^2 = 0.1787**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	-0.01	0.02	-0.04 to 0.02	-0.02	0.02	-0.05 to 0.02
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.02[b]	0.01	0.01 to 0.04	0.02[b]	0.01	0.00 to 0.04
Constant	16.18[a]	0.23	15.74 to 16.62	15.83[a]	0.25	15.34 to 16.31	15.11[a]	0.32	14.49 to 15.74	14.89[a]	0.35	14.19 to 15.58	15.87[a]	1.02	13.87 to 17.87	15.89[a]	1.06	13.81 to 17.96

Note. Significant to $p<.001$[a], $p<.01$[b], $p<.05$[c]

Table 3.19. Linear regression models for Foundation Stage Profile Personal, Social and Emotional Development - poverty

	Model 1 **Expressive vocabulary only** **N = 7,012**						**Model 3** **Expressive vocabulary x poverty** **N = 6,946**						**Model 5** **Expressive vocabulary x poverty with covariates** **N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0794			**R^2 = 0.0851**			**R^2 = 0.0974**			**R^2 = 0.1071**			**R^2 = 0.1710**			**R^2 = 0.1774**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.10[a]	0.00	0.09 to 0.11	0.11[a]	0.00	0.10 to 0.12	0.07[a]	0.01	0.06 to 0.09	0.09[a]	0.01	0.08 to 0.11	0.03[b]	0.01	0.01 to 0.05	0.04[a]	0.01	0.02 to 0.07
Poverty (above threshold)	-	-	-	-	-	-	0.35[c]	0.48	-0.60 to 1.29	1.50[b]	0.54	0.43 to 2.56	0.67	0.58	-0.46 to 1.80	1.39[c]	0.64	0.13 to 2.65
Interaction	-	-	-	-	-	-	0.02	0.01	0.00 to 0.04	0.00	0.01	-0.02 to 0.02	0.00	0.01	-0.02 to 0.02	-0.02	0.01	-0.04 to 0.01
Gender assigned at birth (female)	-	-	-	-	-	-	-	-	-	-	-	-	0.95[a]	0.10	0.76 to 1.14	0.94[a]	0.10	0.74 to 1.15

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0794			**R^2 = 0.0851**			**R^2 = 0.0974**			**R^2 = 0.1071**			**R^2 = 0.1710**			**R^2 = 0.1774**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.09[a]	0.02	-0.13 to -0.05	-0.09[a]	0.02	-0.13 to -0.05
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.05[a]	0.00	0.04 to 0.06	0.05[a]	0.00	0.04 to 0.06
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.08[a]	0.01	-0.10 to -0.06	-0.08[a]	0.01	-0.11 to -0.06
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	0.11	0.27	-0.43 to 0.64	0.14	0.29	-0.43 to 0.70
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	0.46[c]	0.22	0.03 to 0.89	0.51[c]	0.23	0.05 to 0.96

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	$R^2 = 0.0794$			$R^2 = 0.0851$			$R^2 = 0.0974$			$R^2 = 0.1071$			$R^2 = 0.1710$			$R^2 = 0.1774$		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	0.75[a]	0.23	0.29 to 1.21	0.74[a]	0.25	0.25 to 1.23
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	1.07[a]	0.22	0.63 to 1.51	1.07[a]	0.24	0.60 to 1.54
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	1.14[a]	0.31	0.53 to 1.75	0.96[a]	0.33	0.31 to 1.60

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0794			**R^2 = 0.0851**			**R^2 = 0.0974**			**R^2 = 0.1071**			**R^2 = 0.1710**			**R^2 = 0.1774**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.01	-0.02 to 0.02	0.00	0.01	-0.02 to 0.02
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	-0.01	0.02	-0.04 to 0.02	-0.01	0.02	-0.05 to 0.02
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.02[b]	0.01	0.01 to 0.04	0.02[b]	0.01	0.00 to 0.04
Constant	16.18[a]	0.23	15.74 to 16.62	15.83[a]	0.25	15.34 to 16.31	16.70[a]	0.38	15.94 to 17.45	15.56[a]	0.45	14.68 to 16.44	16.62[a]	1.07	14.53 to 18.71	16.08[a]	1.11	13.90 to 18.27

Note. Significant to p<.001[a], p<.01[b], p<.05[c]

Communication, Language and Literacy

Table 3.20. Linear regression models for Foundation Stage Profile Communication, Language and Literacy- gender assigned at birth

	Model 1 **Expressive vocabulary only** **N = 7,012**						**Model 2** **Expressive vocabulary x gender assigned at birth** **N = 7,012**						**Model 4** **Expressive vocabulary x gender assigned at birth with covariates** **N = 5,718**					
	Unadjusted for sample weighting **R^2 = 0.1449**			**Adjusted for sample weighting** **R^2 = 0.1488**			**Unadjusted for sample weighting** **R^2 = 0.1591**			**Adjusted for sample weighting** **R^2 = 0.1607**			**Unadjusted for sample weighting** **R^2 = 0.2860**			**Adjusted for sample weighting** **R^2 = 0.2950**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.23[a]	0.01	0.21 to 0.24	0.24[a]	0.01	0.22 to 0.25	0.24[a]	0.01	0.22 to 0.26	0.25[a]	0.01	0.23 to 0.27	0.08[a]	0.01	0.06 to 0.10	0.08[a]	0.01	0.06 to 0.11
Gender assigned at birth (female)	-	-	-	-	-	-	3.77[a]	0.68	2.43 to 5.10	3.26[a]	0.77	1.75 to 4.78	3.32[a]	0.75	1.85 to 4.80	2.84[a]	0.82	1.24 to 4.44

	Model 1 **Expressive vocabulary only** **N = 7,012**						**Model 2** **Expressive vocabulary x gender assigned at birth** **N = 7,012**						**Model 4** **Expressive vocabulary x gender assigned at birth with covariates** **N = 5,718**					
	Unadjusted for sample weighting **R^2 = 0.1449**			**Adjusted for sample weighting** **R^2 = 0.1488**			**Unadjusted for sample weighting** **R^2 = 0.1591**			**Adjusted for sample weighting** **R^2 = 0.1607**			**Unadjusted for sample weighting** **R^2 = 0.2860**			**Adjusted for sample weighting** **R^2 = 0.2950**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Interaction	-	-	-	-	-	-	-0.05[a]	0.01	-0.07 to -0.02	-0.04[b]	0.01	-0.07 to -0.01	-0.04[a]	0.01	-0.07 to -0.01	-0.03[c]	0.02	-0.06 to 0.00
Poverty (above threshold)	-	-	-	-	-	-	-	-	-	-	-	-	0.97[a]	0.21	0.56 to 1.37	1.01[a]	0.22	0.58 to 1.45
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.16[a]	0.03	-0.22 to -0.10	-0.18[a]	0.03	-0.25 to -0.11
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.14[a]	0.01	0.13 to 0.15	0.14[a]	0.01	0.13 to 0.15

	Model 1 Expressive vocabulary only N = 7,012						Model 2 Expressive vocabulary x gender assigned at birth N = 7,012						Model 4 Expressive vocabulary x gender assigned at birth with covariates N = 5,718					
	Unadjusted for sample weighting $R^2 = 0.1449$			Adjusted for sample weighting $R^2 = 0.1488$			Unadjusted for sample weighting $R^2 = 0.1591$			Adjusted for sample weighting $R^2 = 0.1607$			Unadjusted for sample weighting $R^2 = 0.2860$			Adjusted for sample weighting $R^2 = 0.2950$		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.08[a]	0.02	-0.11 to -0.04	-0.08[a]	0.02	-0.11 to -0.04
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	0.42	0.40	-0.36 to 1.20	0.55[a]	0.42	-0.28 to 1.38
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	1.14[a]	0.33	0.49 to 1.79	1.25[a]	0.36	0.55 to 1.94
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	1.44[a]	0.36	0.73 to 2.15	1.50[a]	0.38	0.75 to 2.26

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1449			**R^2 = 0.1488**			**R^2 = 0.1591**			**R^2 = 0.1607**			**R^2 = 0.2860**			**R^2 = 0.2950**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	2.14[a]	0.35	1.46 to 2.81	2.14[a]	0.37	1.41 to 2.86
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	2.79[a]	0.46	1.90 to 3.68	2.57[a]	0.49	1.61 to 3.53
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.01	0.01	-0.02 to 0.04	0.01	0.02	-0.02 to 0.04

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1449			**R^2 = 0.1488**			**R^2 = 0.1591**			**R^2 = 0.1607**			**R^2 = 0.2860**			**R^2 = 0.2950**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	-0.03	0.02	-0.08 to 0.01	-0.05[c]	0.02	-0.10 to 0.00
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.06[a]	0.01	0.04 to 0.08	0.06[a]	0.01	0.04 to 0.09
Constant	14.33[a]	0.34	13.65 to 15.00	13.72[a]	0.39	12.95 to 14.48	12.81[a]	0.49	11.86 to 13.76	12.50[a]	0.55	11.43 to 13.57	9.66[a]	1.54	6.65 to 12.67	9.75[a]	1.62	6.58 to 12.93

Note. Significant to p<.001[a], p<.01[b], p<.05[c]

Table 3.21. Linear regression models for Foundation Stage Profile Communication, Language and Literacy- poverty

	Model 1						**Model 3**						**Model 5**					
	Expressive vocabulary only						**Expressive vocabulary x poverty**						**Expressive vocabulary x poverty with covariates**					
	N = 7,012						**N = 6,946**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1449			**R^2 = 0.1488**			**R^2 = 0.1705**			**R^2 = 0.1780**			**R^2 = 0.2850**			**R^2 = 0.2950**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.23[a]	0.01	0.21 to 0.24	0.24[a]	0.01	0.22 to 0.25	0.18[a]	0.01	0.16 to 0.21	0.22[a]	0.01	0.19 to 0.25	0.07[a]	0.02	0.04 to - 0.10	0.09[a]	0.02	0.06 to 0.13
Poverty (above threshold)	-	-	-	-	-	-	1.63[c]	0.74	0.18 to 3.07	3.41[a]	0.83	1.78 to 5.03	1.71[c]	0.86	0.02 to 3.40	2.86[a]	0.93	1.02 to 4.69
Interaction	-	-	-	-	-	-	0.02	0.02	-0.01 to 0.05	-0.01	0.02	-0.05 to 0.02	-0.02	0.02	-0.05 to 0.02	-0.04[c]	0.02	-0.07 to 0.00
Gender assigned at birth (female)	-	-	-	-	-	-	-	-	-	-	-	-	1.21[a]	0.15	0.92 to 1.51	1.17[a]	0.16	0.86 to 1.48

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1449			**R^2 = 0.1488**			**R^2 = 0.1705**			**R^2 = 0.1780**			**R^2 = 0.2850**			**R^2 = 0.2950**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.16[a]	0.03	-0.22 to -0.10	-0.18[a]	0.03	-0.25 to -0.11
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.14[a]	0.01	0.13 to 0.15	0.14[a]	0.01	0.13 to 0.15
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.08[a]	0.02	-0.11 to -0.04	-0.08[a]	0.02	-0.11 to -0.04
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	0.39	0.40	-0.40 to 1.17	0.51	0.42	-0.32 to 1.34
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	1.12[a]	0.33	0.46 to 1.77	1.21[a]	0.36	0.51 to 1.91

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1449			**R^2 = 0.1488**			**R^2 = 0.1705**			**R^2 = 0.1780**			**R^2 = 0.2850**			**R^2 = 0.2950**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	1.41[a]	0.36	0.70 to 2.13	1.46[a]	0.39	0.70 to 2.21
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	2.12[a]	0.35	1.45 to 2.80	2.12[a]	0.37	1.39 to 2.84
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	2.78[a]	0.46	1.89 to 3.67	2.54[a]	0.49	1.58 to 3.50

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1449			**R^2 = 0.1488**			**R^2 = 0.1705**			**R^2 = 0.1780**			**R^2 = 0.2850**			**R^2 = 0.2950**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.01	0.01	-0.02 to 0.04	0.01	0.02	-0.02 to 0.04
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	-0.03	0.02	-0.08 to 0.02	-0.05[c]	0.02	-0.10 to 0.00
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.06[a]	0.01	0.04 to 0.08	0.06[a]	0.01	0.04 to 0.09
Constant	14.33[a]	0.34	13.65 to 15.00	13.72[a]	0.39	12.95 to 14.48	14.68[a]	0.57	13.56 to 15.81	12.80[a]	0.66	11.51 to 14.09	10.11[a]	1.61	6.96 to 13.27	9.21[a]	1.70	5.88 to 12.55

Note. Significant to p<.001[a], p<.01[b], p<.05[c]

Mathematical development

Table 3.22. Linear regression models for Foundation Stage Profile Mathematical development- gender assigned at birth

	Model 1						**Model 2**						**Model 4**					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1274			**R^2 = 0.1280**			**R^2 = 0.1303**			**R^2 = 0.1297**			**R^2 = 0.2408**			**R^2 = 0.2462**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.14[a]	0.01	0.13 to 0.15	0.14[a]	0.01	0.13 to 0.15	0.16[a]	0.01	0.16	0.16[a]	0.01	0.14 to 0.17	0.05[a]	0.01	0.03 to 0.06	0.05[a]	0.01	0.04 to 0.07
Gender assigned at birth (female)	-	-	-	-	-	-	2.05[a]	0.49	2.05	1.70[a]	0.54	0.64 to 2.76	1.66[a]	0.53	0.63 to 2.70	1.38[b]	0.56	0.29 to 2.48
Interaction	-	-	-	-	-	-	-0.04[a]	0.01	-0.04	-0.03[a]	0.01	-0.05 to -0.01	-0.03[a]	0.01	-0.05 to -0.01	-0.03[b]	0.01	-0.05 to -0.01

	Model 1						**Model 2**						**Model 4**					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1274			**R^2 = 0.1280**			**R^2 = 0.1303**			**R^2 = 0.1297**			**R^2 = 0.2408**			**R^2 = 0.2462**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Poverty (above threshold)	-	-	-	-	-	-	-	-	-	-	-	-	0.41[a]	0.14	0.13 to 0.68	0.38[b]	0.15	0.09 to 0.67
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.09[a]	0.02	-0.13 to -0.05	-0.10[a]	0.02	-0.15 to -0.06
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.09[a]	0.00	0.08 to 0.10	0.09[a]	0.00	0.08 to 0.10
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.05[a]	0.01	-0.07 to -0.03	-0.05[a]	0.01	-0.07 to -0.03

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1274			**R^2 = 0.1280**			**R^2 = 0.1303**			**R^2 = 0.1297**			**R^2 = 0.2408**			**R^2 = 0.2462**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	0.38	0.28	-0.17 to 0.93	0.38	0.30	-0.20 to 0.96
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	0.86[a]	0.24	0.40 to 1.32	0.93[a]	0.25	0.44 to 1.42
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	1.14[a]	0.25	0.64 to 1.63	1.18[a]	0.27	0.66 to 1.70

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1274			**R^2 = 0.1280**			**R^2 = 0.1303**			**R^2 = 0.1297**			**R^2 = 0.2408**			**R^2 = 0.2462**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	1.44[a]	0.24	0.97 to 1.92	1.46[a]	0.26	0.96 to 1.96
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	1.74[a]	0.30	1.15 to 2.33	1.68[a]	0.31	1.07 to 2.30
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.01	-0.02 to 0.02	0.00	0.01	-0.02 to 0.02
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	-0.01	0.02	-0.04 to 0.02	-0.02	0.02	-0.06 to 0.01

	Model 1						**Model 2**						**Model 4**					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1274			**R^2 = 0.1280**			**R^2 = 0.1303**			**R^2 = 0.1297**			**R^2 = 0.2408**			**R^2 = 0.2462**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.02[c]	0.01	0.00 to 0.03	0.02[a]	0.01	0.00 to 0.04
Constant	13.73[a]	0.25	13.25 to 14.21	13.47[a]	0.27	12.94 to 14.01	12.78[a]	0.35	12.78	12.70[a]	0.39	11.94 to 13.45	10.76[a]	1.03	8.74 to 12.78	11.15[a]	1.08	9.02 to 13.27

Note. Significant to p<.001[a], p<.01[b], p<.05[c]

Table 3.23. Linear regression models for Foundation Stage Profile Mathematical development- poverty

	Model 1						**Model 3**						**Model 5**					
	Expressive vocabulary only						**Expressive vocabulary x poverty**						**Expressive vocabulary x poverty with covariates**					
	N = 7,012						**N = 6,946**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1274			**R^2 = 0.1280**			**R^2 = 0.1448**			**R^2 = 0.1485**			**R^2 = 0.2396**			**R^2 = 0.2459**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.14[a]	0.01	0.13 to 0.15	0.14[a]	0.01	0.13 to 0.15	0.12 [a]	0.01	0.10 to 0.14	0.14[a]	0.01	0.12 to 0.16	0.05[a]	0.01	0.02 to 0.07	0.06[a]	0.01	0.04 to 0.08
Poverty (above threshold)	-	-	-	-	-	-	1.29 [c]	0.53	0.25 to 2.32	2.26[a]	0.58	1.11 to 3.41	1.23[c]	0.62	0.03 to 2.44	1.65[b]	0.64	0.39 to 2.91
Interaction	-	-	-	-	-	-	0.00	0.01	-0.02 to 0.02	-0.02	0.01	-0.04 to 0.01	-0.02	0.01	-0.04 to 0.01	-0.03[c]	0.01	-0.05 to 0.00
Gender assigned at birth (female)	-	-	-	-	-	-	-	-	-	-	-	-	0.05	0.10	-0.15 to 0.24	0.02	0.10	-0.19 to 0.22

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1274			**R^2 = 0.1280**			**R^2 = 0.1448**			**R^2 = 0.1485**			**R^2 = 0.2396**			**R^2 = 0.2459**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.09[a]	0.02	-0.13 to -0.05	-0.10[a]	0.02	-0.15 to -0.06
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.09[a]	0.00	0.08 to 0.10	0.09[a]	0.00	0.08 to 0.10
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.05[a]	0.01	-0.07 to -0.03	-0.05[a]	0.01	-0.07 to -0.03
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	0.35	0.28	-0.20 to 0.90	0.35	0.30	-0.24 to 0.93
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	0.84[a]	0.24	0.37 to 1.30	0.90[a]	0.25	0.41 to 1.39

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	$R^2 = 0.1274$			$R^2 = 0.1280$			$R^2 = 0.1448$			$R^2 = 0.1485$			$R^2 = 0.2396$			$R^2 = 0.2459$		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	1.11[a]	0.25	0.62 to 1.61	1.15[a]	0.27	0.62 to 1.67
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	1.43[a]	0.24	0.95 to 1.91	1.45[a]	0.26	0.95 to 1.95
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	1.73[a]	0.30	1.14 to 2.32	1.66[a]	0.31	1.05 to 2.28

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.1274			**R^2 = 0.1280**			**R^2 = 0.1448**			**R^2 = 0.1485**			**R^2 = 0.2396**			**R^2 = 0.2459**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.01	-0.02 to 0.02	0.00	0.01	-0.02 to 0.02
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	-0.01	0.02	-0.04 to 0.02	-0.02	0.02	-0.06 to 0.01
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.02[c]	0.01	0.00 to 0.03	0.02[c]	0.01	0.00 to 0.03
Constant	13.73[a]	0.25	13.25 to 14.21	13.47[a]	0.27	12.94 to 14.01	13.75[a]	0.42	12.93 to 14.57	12.76[a]	0.47	11.84 to 13.68	10.93[a]	1.09	8.79 to 13.07	10.87[a]	1.15	8.62 to 13.12

Note. Significant to p<.001[a], p<.01[b], p<.05[c]

Knowledge and understanding of the world

Table 3.24. Linear regression models for Foundation Stage Profile Knowledge and understanding of the world- gender assigned at birth

	Model 1						**Model 2**						**Model 4**					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0882			**R^2 = 0.0910**			**R^2 = 0.0887**			**R^2 = 0.0914**			**R^2 = 0.1491**			**R^2 = 0.1572**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.04[a]	0.00	0.04 to 0.04	0.04[a]	0.00	0.04 to 0.05	0.04[a]	0.00	0.04 to 0.05	0.04[a]	0.00	0.04	0.02[a]	0.00	0.01 to 0.02	0.02[a]	0.00	0.01 to 0.02
Gender assigned at birth (female)	-	-	-	-	-	-	0.21	0.17	-0.13 to 0.54	0.12	0.19	0.12	0.26[c]	0.19	-0.11 to 0.64	0.15	0.21	-0.26 to 0.57

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0882			**R^2 = 0.0910**			**R^2 = 0.0887**			**R^2 = 0.0914**			**R^2 = 0.1491**			**R^2 = 0.1572**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Interaction	-	-	-	-	-	-	-0.01	0.00	-0.01 to 0.00	0.00	0.00	0.00	-0.01	0.00	-0.01 to 0.00	0.00	0.00	-0.01 to 0.00
Poverty (above threshold)	-	-	-	-	-	-	-	-	-	-	-	-	0.27[a]	0.05	0.16 to 0.37	0.29[a]	0.06	0.17 to 0.40
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.03[a]	0.01	-0.04 to -0.01	-0.03[a]	0.01	-0.05 to -0.01
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.02[a]	0.00	0.02 to 0.02	0.02[a]	0.00	0.02 to 0.02

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	$R^2 = 0.0882$			**$R^2 = 0.0910$**			**$R^2 = 0.0887$**			**$R^2 = 0.0914$**			**$R^2 = 0.1491$**			**$R^2 = 0.1572$**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.02[a]	0.00	-0.03 to -0.01	-0.02[a]	0.00	-0.03 to -0.01
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	0.08	0.10	-0.12 to 0.29	0.05	0.11	-0.17 to 0.28
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	0.25[a]	0.08	0.08 to 0.41	0.27[a]	0.09	0.10 to 0.45
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	0.35[a]	0.09	0.18 to 0.53	0.34[a]	0.10	0.15 to 0.53

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						Expressive vocabulary x gender assigned at birth						Expressive vocabulary x gender assigned at birth with covariates					
	N = 7,012						N = 7,012						N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	$R^2 = 0.0882$			$R^2 = 0.0910$			$R^2 = 0.0887$			$R^2 = 0.0914$			$R^2 = 0.1491$			$R^2 = 0.1572$		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	0.44[a]	0.09	0.27 to 0.61	0.46[a]	0.09	0.28 to 0.64
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	0.59[a]	0.11	0.36 to 0.81	0.57[a]	0.12	0.33 to 0.81
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.00	-0.01 to 0.01	0.00	0.00	-0.01 to 0.00

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						Expressive vocabulary x gender assigned at birth						Expressive vocabulary x gender assigned at birth with covariates					
	N = 7,012						N = 7,012						N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	$R^2 = 0.0882$			$R^2 = 0.0910$			$R^2 = 0.0887$			$R^2 = 0.0914$			$R^2 = 0.1491$			$R^2 = 0.1572$		
Measure	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.01	-0.01 to 0.01	0.00	0.01	-0.01 to 0.01
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.01[b]	0.00	0.00 to 0.01	0.01[b]	0.00	0.00 to 0.01
Constant	4.76[a]	0.09	4.59 to 4.93	4.68[a]	0.10	4.49 to 4.86	4.65[a]	0.12	4.41 to 4.89	4.60[a]	0.14	4.60	4.47[a]	0.39	3.70 to 5.24	4.61[a]	0.41	3.81 to 5.41

Note. Significant to p<.001[a], p<.01[b], p<.05[c]

Table 3.25. Linear regression models for Foundation Stage Profile Knowledge and understanding of the world- poverty

	Model 1						**Model 3**						**Model 5**					
	Expressive vocabulary only						**Expressive vocabulary x poverty**						**Expressive vocabulary x poverty with covariates**					
	N = 7,012						**N = 6,946**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0882			**R^2 = 0.0910**			**R^2 = 0.1063**			**R^2 = 0.1126**			**R^2 = 0.1487**			**R^2 = 0.1574**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.04[a]	0.00	0.04 to 0.04	0.04[a]	0.00	0.04 to 0.05	0.03[a]	0.00	0.03 to 0.04	0.04[a]	0.00	0.03 to 0.05	0.02[a]	0.00	0.01 to 0.03	0.02[a]	0.00	0.01 to 0.03
Poverty (above threshold)	-	-	-	-	-	-	0.33	0.19	-0.03 to 0.70	0.69[a]	0.21	0.28 to 1.09	0.48[c]	0.22	0.04 to 0.92	0.64[b]	0.25	0.15 to 1.12
Interaction	-	-	-	-	-	-	0.00	0.00	-0.00 to 0.01	0.00	0.00	-0.01 to 0.00	0.00	0.00	-0.01 to 0.00	-0.01	0.00	-0.02 to 0.00
Gender assigned at birth (female)	-	-	-	-	-	-	-	-	-	-	-	-	-0.08[a]	0.04	-0.16 to -0.01	-0.09[c]	0.04	-0.16 to -0.01

	Model 1						Model 3						Model 5					
	Expressive vocabulary only						Expressive vocabulary x poverty						Expressive vocabulary x poverty with covariates					
	N = 7,012						N = 6,946						N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	R^2 = 0.0882			R^2 = 0.0910			R^2 = 0.1063			R^2 = 0.1126			R^2 = 0.1487			R^2 = 0.1574		
Measure	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.03[c]	0.01	-0.04 to -0.01	-0.03[a]	0.01	-0.05 to -0.01
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.02[a]	0.00	0.02 to 0.02	0.02[a]	0.00	0.02 to 0.02
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.02[a]	0.00	-0.03 to -0.01	-0.02[a]	0.00	-0.03 to -0.01
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	0.08	0.10	-0.13 to 0.28	0.05	0.12	-0.18 to 0.27
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	0.24[a]	0.08	0.08 to 0.41	0.27[a]	0.09	0.09 to 0.44

	Model 1						Model 3						Model 5					
	Expressive vocabulary only						**Expressive vocabulary x poverty**						**Expressive vocabulary x poverty with covariates**					
	N = 7,012						**N = 6,946**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0882			**R^2 = 0.0910**			**R^2 = 0.1063**			**R^2 = 0.1126**			**R^2 = 0.1487**			**R^2 = 0.1574**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	0.34[a]	0.09	0.17 to 0.52	0.33[a]	0.10	0.14 to 0.52
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	0.44[a]	0.09	0.27 to 0.61	0.46[a]	0.09	0.27 to 0.64
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	0.58[a]	0.11	0.36 to 0.80	0.57[a]	0.12	0.33 to 0.81

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	$R^2 = 0.0882$			$R^2 = 0.0910$			$R^2 = 0.1063$			$R^2 = 0.1126$			$R^2 = 0.1487$			$R^2 = 0.1574$		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.00	-0.01 to 0.01	0.00	0.00	-0.01 to 0.00
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.01	-0.01 to 0.01	0.00	0.01	-0.01 to 0.01
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.01[b]	0.00	0.00 to 0.01	0.01[b]	0.00	0.00 to 0.01
Constant	4.76[a]	0.09	4.59 to 4.93	4.68[a]	0.10	4.49 to 4.86	4.84[a]	0.15	4.55 to 5.13	4.50[a]	0.17	4.16 to 4.83	4.48[a]	0.41	3.68 to 5.29	4.47[a]	0.43	3.62 to 5.32

Note. Significant to p<.001[a], p<.01[b], p<.05[c]

Physical development

Table 3.26. Linear regression models for Foundation Stage Profile Physical development- gender assigned at birth

	Model 1						**Model 2**						**Model 4**					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0573			**R^2 = 0.0633**			**R^2 = 0.0713**			**R^2 = 0.0780**			**R^2 = 0.1166**			**R^2 = 0.1214**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.03[a]	0.00	0.03 to 0.03	0.03[a]	0.00	0.03 to 0.03	0.03[a]	0.00	0.03 to 0.04	0.03[a]	0.00	0.03 to 0.04	0.01[a]	0.00	0.01 to 0.02	0.02[a]	0.00	0.01 to 0.02
Gender assigned at birth (female)	-	-	-	-	-	-	0.74[a]	0.15	0.44 to 1.03	0.78[a]	0.16	0.46 to 1.09	0.68[a]	0.17	0.35 to 1.02	0.72[a]	0.18	0.36 to 1.08
Interaction	-	-	-	-	-	-	-0.01[b]	0.00	-0.01 to -0.00	-0.01[a]	0.00	-0.02 to 0.00	-0.01[a]	0.00	-0.01 to 0.00	-0.01[b]	0.00	-0.02 to 0.00

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						Expressive vocabulary x gender assigned at birth						Expressive vocabulary x gender assigned at birth with covariates					
	N = 7,012						N = 7,012						N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	$R^2 = 0.0573$			$R^2 = 0.0633$			$R^2 = 0.0713$			$R^2 = 0.0780$			$R^2 = 0.1166$			$R^2 = 0.1214$		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Poverty (above threshold)	-	-	-	-	-	-	-	-	-	-	-	-	0.19[a]	0.05	0.09 to 0.28	0.20[a]	0.05	0.10 to 0.30
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.03[a]	0.01	-0.04 to -0.02	-0.03[a]	0.01	-0.04 to -0.01
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.01[a]	0.00	0.01 to 0.02	0.01[a]	0.00	0.01 to 0.02
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.02[a]	0.00	-0.03 to -0.01	-0.02	0.00	-0.03 to -0.01

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0573			**R^2 = 0.0633**			**R^2 = 0.0713**			**R^2 = 0.0780**			**R^2 = 0.1166**			**R^2 = 0.1214**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	-0.03	0.09	-0.22 to 0.15	-0.06	0.10	-0.26 to 0.15
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	0.10	0.08	-0.05 to 0.25	0.09	0.08	-0.07 to 0.25
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	0.14	0.08	-0.02 to 0.30	0.13	0.09	-0.04 to 0.30

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0573			**R^2 = 0.0633**			**R^2 = 0.0713**			**R^2 = 0.0780**			**R^2 = 0.1166**			**R^2 = 0.1214**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	0.25[a]	0.08	0.10 to 0.40	0.23[b]	0.08	0.07 to 0.39
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	0.26[b]	0.10	0.05 to 0.46	0.22[c]	0.11	0.01 to 0.43
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.00	-0.01 to 0.00	-0.01	0.00	-0.01 to 0.00
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.01	-0.01 to 0.01	0.00	0.01	-0.01 to 0.01

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0573			**R^2 = 0.0633**			**R^2 = 0.0713**			**R^2 = 0.0780**			**R^2 = 0.1166**			**R^2 = 0.1214**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.00	0.00 to 0.01	0.00	0.00	0.00 to 0.01
Constant	5.88[a]	0.08	5.74 to 6.03	5.77[a]	0.08	5.60 to 5.93	5.59[a]	0.11	5.38 to 5.80	5.47[a]	0.12	5.24 to 5.71	6.23[a]	0.35	5.55 to 6.92	6.10[a]	0.35	5.42 to 6.79

Note. Significant to p<.001[a], p<.01[b], p<.05[c]

Table 3.27. Linear regression models for Foundation Stage Profile Physical development- poverty

	Model 1 **Expressive vocabulary only** **N = 7,012**						**Model 3** **Expressive vocabulary x poverty** **N = 6,946**						**Model 5** **Expressive vocabulary x poverty with covariates** **N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0573			**R^2 = 0.0633**			**R^2 = 0.0689**			**R^2 = 0.0776**			**R^2 = 0.1158**			**R^2 = 0.1213**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.03[a]	0.00	0.03 to 0.03	0.03[a]	0.00	0.03 to 0.03	0.02[a]	0.00	0.02 to 0.03	0.03[a]	0.00	0.02 to 0.04	0.01[a]	0.00	0.01 to 0.02	0.02[a]	0.00	0.01 to 0.03
Poverty (above threshold)	-	-	-	-	-	-	0.20	0.16	-0.12 to 0.52	0.56[a]	0.18	0.21 to 0.92	0.43[c]	0.20	0.03 to 0.82	0.65[a]	0.22	0.22 to 1.08
Interaction	-	-	-	-	-	-	0.00	0.00	-0.00 to 0.01	0.00	0.00	-0.01 to 0.00	-0.01	0.00	-0.01 to 0.00	-0.01[c]	0.00	-0.02 to 0.00
Gender assigned at birth (female)	-	-	-	-	-	-	-	-	-	-	-	-	0.28[a]	0.03	0.21 to 0.34	0.29[a]	0.04	0.22 to 0.35

	Model 1						Model 3						Model 5					
	Expressive vocabulary only						**Expressive vocabulary x poverty**						**Expressive vocabulary x poverty with covariates**					
	N = 7,012						**N = 6,946**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0573			**R^2 = 0.0633**			**R^2 = 0.0689**			**R^2 = 0.0776**			**R^2 = 0.1158**			**R^2 = 0.1213**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.03[a]	0.01	-0.04 to -0.02	-0.03[a]	0.01	-0.04 to -0.01
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.01[a]	0.00	0.01 to 0.02	0.01[a]	0.00	0.01 to 0.02
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.02[a]	0.00	-0.03 to -0.01	-0.02[a]	0.00	-0.03 to -0.01
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	-0.04	0.09	-0.23 to 0.14	-0.07	0.10	-0.27 to 0.13
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	0.09	0.08	-0.06 to 0.24	0.08	0.08	-0.08 to 0.24

Measure	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0573			**R^2 = 0.0633**			**R^2 = 0.0689**			**R^2 = 0.0776**			**R^2 = 0.1158**			**R^2 = 0.1213**		
	Coef.	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	0.13	0.08	-0.03 to 0.29	0.12	0.09	-0.06 to 0.29
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	0.24[a]	0.08	0.09 to 0.40	0.23[b]	0.08	0.06 to 0.39
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	0.25[c]	0.10	0.05 to 0.46	0.21[c]	0.11	0.00 to 0.43

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0573			**R^2 = 0.0633**			**R^2 = 0.0689**			**R^2 = 0.0776**			**R^2 = 0.1158**			**R^2 = 0.1213**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.00	-0.01 to 0.00	-0.01	0.00	-0.01 to 0.00
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.01	-0.01 to 0.01	0.00	0.01	-0.01 to 0.01
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.00	0.00 to 0.01	0.00	0.00	0.00 to 0.01
Constant	5.88[a]	0.08	5.74 to 6.03	5.77[a]	0.08	5.60 to 5.93	5.95[a]	0.13	5.69 to 6.20	5.58[a]	0.15	5.29 to 5.87	6.25[a]	0.37	5.53 to 6.97	5.99[a]	0.37	5.26 to 6.72

Note. Significant to p<.001[a], p<.01[b], p<.05[c]

Creative development

Table 3.28. Linear regression models for Foundation Stage Profile Creative development- gender assigned at birth

	Model 1						**Model 2**						**Model 4**					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0733			**R^2 = 0.0770**			**R^2 = 0.1077**			**R^2 = 0.1111**			**R^2 = 0.1570**			**R^2 = 0.1624**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.04[a]	0.00	0.03 to 0.04	0.04[a]	0.00	0.03 to 0.04	0.04[a]	0.00	0.03 to 0.04	0.04[a]	0.00	0.03 to 0.04	0.02[a]	0.00	0.01 to 0.02	0.02[a]	0.00	0.01 to 0.02
Gender assigned at birth (female)	-	-	-	-	-	-	0.89[a]	0.15	0.59 to 1.19	0.89[a]	0.17	0.56 to 1.22	1.01[a]	0.18	0.66 to 1.35	0.91[a]	0.19	0.53 to 1.29
Interaction	-	-	-	-	-	-	-0.01[c]	0.00	-0.01 to 0.00	-0.01[c]	0.00	-0.01 to 0.00	-0.01[a]	0.00	-0.02 to 0.00	-0.01[c]	0.00	-0.01 to 0.00

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0733			**R^2 = 0.0770**			**R^2 = 0.1077**			**R^2 = 0.1111**			**R^2 = 0.1570**			**R^2 = 0.1624**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Poverty (above threshold)	-	-	-	-	-	-	-	-	-	-	-	-	0.21[a]	0.05	0.11 to 0.30	0.20[a]	0.05	0.09 to 0.30
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.02[c]	0.01	-0.03 to 0.00	-0.02[c]	0.01	-0.03 to 0.00
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.01[a]	0.00	0.01 to 0.02	0.01[a]	0.00	0.01 to 0.02
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.02[a]	0.00	-0.02 to -0.01	-0.02[a]	0.00	-0.03 to -0.01

	Model 1						**Model 2**						**Model 4**					
	Expressive vocabulary only						**Expressive vocabulary x gender assigned at birth**						**Expressive vocabulary x gender assigned at birth with covariates**					
	N = 7,012						**N = 7,012**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0733			**R^2 = 0.0770**			**R^2 = 0.1077**			**R^2 = 0.1111**			**R^2 = 0.1570**			**R^2 = 0.1624**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	0.14	0.09	-0.05 to 0.32	0.13	0.10	-0.07 to 0.33
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	0.21[b]	0.08	0.06 to 0.37	0.23[b]	0.08	0.07 to 0.40
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	0.36[a]	0.08	0.19 to 0.52	0.36[a]	0.09	0.19 to 0.54

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						Expressive vocabulary x gender assigned at birth						Expressive vocabulary x gender assigned at birth with covariates					
	N = 7,012						N = 7,012						N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	R^2 = 0.0733			R^2 = 0.0770			R^2 = 0.1077			R^2 = 0.1111			R^2 = 0.1570			R^2 = 0.1624		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	0.45[a]	0.08	0.29 to 0.61	0.47[a]	0.09	0.30 to 0.64
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	0.52[a]	0.11	0.31 to 0.72	0.52[a]	0.11	0.29 to 0.74
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.00	-0.01 to 0.01	0.00	0.00	-0.01 to 0.01
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.01	-0.01 to 0.01	0.00	0.01	-0.01 to 0.01

	Model 1						Model 2						Model 4					
	Expressive vocabulary only						Expressive vocabulary x gender assigned at birth						Expressive vocabulary x gender assigned at birth with covariates					
	N = 7,012						N = 7,012						N = 5,718					
	Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting			Unadjusted for sample weighting			Adjusted for sample weighting		
	$R^2 = 0.0733$			$R^2 = 0.0770$			$R^2 = 0.1077$			$R^2 = 0.1111$			$R^2 = 0.1570$			$R^2 = 0.1624$		
Measure	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI	Coef.	SE	95% CI
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.01[c]	0.00	0.00 to 0.01	0.01	0.00	0.00 to 0.01
Constant	5.03[a]	0.08	4.87 to 5.18	4.93[a]	0.09	4.76 to 5.10	4.70[a]	0.11	4.48 to 4.92	4.64[a]	0.13	4.39 to 4.88	4.43[a]	0.36	3.72 to 5.14	4.52[a]	0.38	3.78 to 5.25

Note. Significant to p<.001[a], p<.01[b], p<.05[c]

Table 3.29. Linear regression models for Foundation Stage Profile Creative development- poverty

	Model 1						**Model 3**						**Model 5**					
	Expressive vocabulary only						**Expressive vocabulary x poverty**						**Expressive vocabulary x poverty with covariates**					
	N = 7,012						**N = 6,946**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0733			**R^2 = 0.0770**			**R^2 = 0.0682**			**R^2 = 0.0916**			**R^2 = 0.1561**			**R^2 = 0.1622**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
BAS-2 Naming Vocabulary	0.04[a]	0.00	0.03 to 0.04	0.04[a]	0.00	0.03 to 0.04	0.03[a]	0.00	0.02 to 0.03	0.03[a]	0.00	0.03 to 0.04	0.02[a]	0.00	0.01 to 0.02	0.02[a]	0.00	0.01 to 0.03
Poverty (above threshold)	-	-	-	-	-	-	0.24	0.17	-0.08 to 0.57	0.55	0.19	0.18 to 0.92	0.48[c]	0.20	0.09 to 0.88	0.59[a]	0.22	0.15 to 1.03
Interaction	-	-	-	-	-	-	0.00	0.00	0.00 to 0.01	0.00[a]	0.00	-0.01 to 0.00	-0.01	0.00	-0.01 to 0.00	-0.01	0.00	-0.02 to 0.00
Gender assigned at birth (female)	-	-	-	-	-	-	-	-	-	-	-	-	0.52[a]	0.04	0.45 to 0.59	0.52[a]	0.04	0.44 to 0.59

	Model 1						Model 3						Model 5					
	Expressive vocabulary only						**Expressive vocabulary x poverty**						**Expressive vocabulary x poverty with covariates**					
	N = 7,012						**N = 6,946**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0733			**R^2 = 0.0770**			**R^2 = 0.0682**			**R^2 = 0.0916**			**R^2 = 0.1561**			**R^2 = 0.1622**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Child age	-	-	-	-	-	-	-	-	-	-	-	-	-0.02[c]	0.01	-0.03 to 0.00	-0.02[c]	0.01	-0.03 to 0.00
Initial school readiness	-	-	-	-	-	-	-	-	-	-	-	-	0.01[a]	0.00	0.01 to 0.02	0.01[a]	0.00	0.01 to 0.02
Behavioural difficulties	-	-	-	-	-	-	-	-	-	-	-	-	-0.02[a]	0.00	-0.02 to -0.01	-0.02[a]	0.00	-0.03 to 0.00
Maternal education (NVQ 1)	-	-	-	-	-	-	-	-	-	-	-	-	0.13	0.09	-0.06 to 0.31	0.12	0.10	-0.08 to 0.32
Maternal education (NVQ 2)	-	-	-	-	-	-	-	-	-	-	-	-	0.21[b]	0.08	0.05 to 0.36	0.22[b]	0.08	0.06 to 0.39

	Model 1 Expressive vocabulary only N = 7,012						Model 3 Expressive vocabulary x poverty N = 6,946						Model 5 Expressive vocabulary x poverty with covariates N = 5,718					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0733			**R^2 = 0.0770**			**R^2 = 0.0682**			**R^2 = 0.0916**			**R^2 = 0.1561**			**R^2 = 0.1622**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal education (NVQ 3)	-	-	-	-	-	-	-	-	-	-	-	-	0.35[a]	0.08	0.19 to 0.52	0.35[a]	0.09	0.18 to 0.53
Maternal education (NVQ 4)	-	-	-	-	-	-	-	-	-	-	-	-	0.44[a]	0.08	0.28 to 0.60	0.47[a]	0.09	0.29 to 0.64
Maternal education (NVQ 5)	-	-	-	-	-	-	-	-	-	-	-	-	0.51[a]	0.11	0.30 to 0.72	0.51[a]	0.11	0.29 to 0.73

	Model 1						**Model 3**						**Model 5**					
	Expressive vocabulary only						**Expressive vocabulary x poverty**						**Expressive vocabulary x poverty with covariates**					
	N = 7,012						**N = 6,946**						**N = 5,718**					
	Unadjusted for sample weighting			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**			**Unadjusted for sample weighting**			**Adjusted for sample weighting**		
	R^2 = 0.0733			**R^2 = 0.0770**			**R^2 = 0.0682**			**R^2 = 0.0916**			**R^2 = 0.1561**			**R^2 = 0.1622**		
Measure	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**	**Coef.**	**SE**	**95% CI**
Maternal age at birth	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.00	-0.01 to 0.01	0.00	0.00	-0.01 to 0.01
Maternal mental health	-	-	-	-	-	-	-	-	-	-	-	-	0.00	0.01	-0.01 to 0.01	0.00	0.01	-0.01 to 0.01
Home Learning Environment	-	-	-	-	-	-	-	-	-	-	-	-	0.01[c]	0.00	0.00 to 0.01	0.01	0.00	0.00 to 0.01
Constant	5.03[a]	0.08	4.87 to 5.18	4.93[a]	0.09	4.76 to 5.10	5.09[a]	0.13	4.83 to 5.35	4.78[a]	0.15	4.47 to 5.08	4.46[a]	0.38	3.73 to 5.20	4.42[a]	0.39	3.65 to 5.19

Note. Significant to p<.001[a], p<.01[b], p<.05[c]

3.5.5. Appendix H. Extreme outliers

Outliers' effects will only be reported for the final models (4 and 5). This is because these are the main models of interest, and models 1-3 are for comparison purposes. Figures 3.23 and 3.24 show box plot outliers. Box plot outliers were the top two scores for BAS-2 NV (i.e. 79 and 80; n=59 in final models) and scores of 40 and under for FSP total (n=69 in final models). Removal of outlier FSP total and/or BAS-2 did not make any significant changes in model 4 (gender assigned at birth as moderator). For model 5, poverty as an individual predictor became non-significant when both types of outliers were removed, or when FSP total outliers were removed alone. As such, there may be some link between poorer FSP total scores and relative poverty. Table 3.30 provides means of the outlier groups by predictor, outcome, moderators and covariates. For the BAS-2 NV outlier group (i.e. very high scores), they showed higher school readiness scores and were characterised by being proportionately more in the hypothesised advantaged individual and social scores and groups. The opposite was generally the case with those in the FSP total outlier group (i.e. very low scores). Very different demographics in each group would suggest why a combination of outlier groups would not change the regression models. Further, outlier groups identified by box plots are disproportionately represented in specific subgroups, with very high expressive vocabulary scorers being linked to more advantage and better scores, while the opposite being the case for very low school readiness scorers.

Outliers via cook's distance were identified when using the 4/n threshold by Hahs-Vaughn and Lomax (2020). 338 outliers were identified for model 4, and 347 for model 5. Visual inspection of the data showed many of the children had lower scores in both expressive vocabulary and school readiness total, or low scoring in one and high scoring in the other. Furthermore, when examining the means and proportions of each variable (table 3.8) found lower averages than the whole (and full case) sample, scoring low on expressive vocabulary and school readiness scores as well as being more represented in male and poverty subgroups. When removing outliers from the final models, this changed many of the variables to non-significant, with only poverty, age, maternal education (NVQ 2 and 4 levels) staying significant for model 4; and age and maternal education (NVQ 2 level) staying significant for model 5.

Figure 3.23. Outliers according to box plot for BAS-2 Naming Vocabulary (scores of 79 and 80, n=7,012)

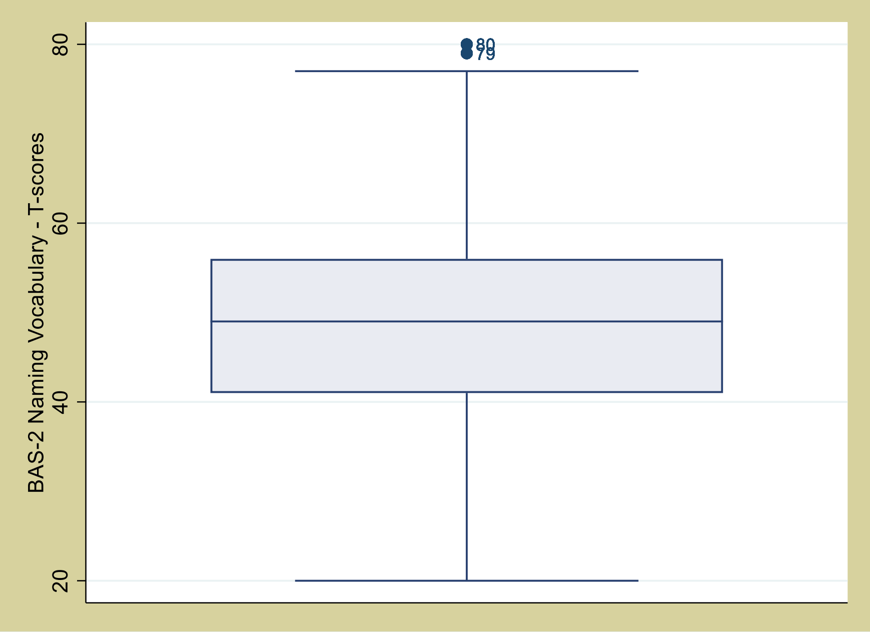
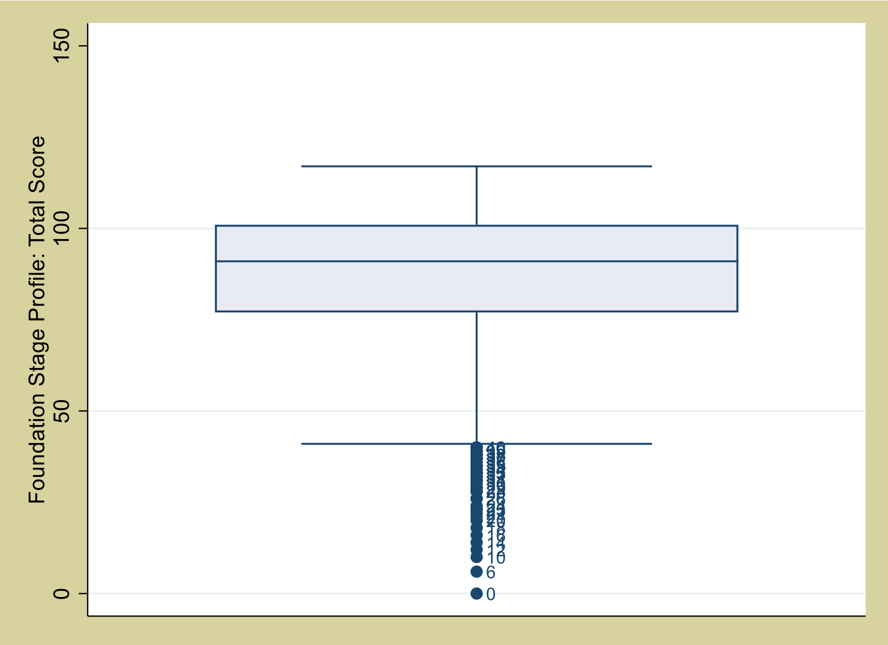


Figure 3.24. Outliers according to box plot for FSP total (scores of 40 and below, n=7,012)



For the logistic regression, least likelihood estimates were used. Ten outlier children were identified. Six achieved GLD and got low BAS-2 scores (at least -1SD). Five of these six children were female, and all lived above the 60% poverty median. Three did not achieve GLD and had below average initial school readiness scores (at least -1SD), but got above average BAS-2 NV scores (at least +1SD). They were all male and living below the 60% median poverty threshold. The last outlier child had below average BAS-2 NV, and did not achieve GLD plus had a below average initial school readiness score (-1SD), but were female and living above 60% median. Outliers therefore seemed to reflect doing well in language, and not doing well in school readiness or vice versa. Additionally, the majority of outliers seemed to be children doing poorly (in at least language initially) despite being female and above poverty; or being male and in poverty and doing well initially at oral language, but then poorly in school readiness. Comparisons with the children removed in models 4 and 5 demonstrated no real differences from the original models.

Overall, there were interesting trends in outlier children’s characteristics, and removal of them seemed to influence results considerably for the linear models. However, it was decided that no outliers should be removed. This was because many of these children were within the lower spectrum of multiple variables, which are an important part of the sample and reflect the spectrum of experiences in a representative sample. The patterns themselves also represent children which are unusual (e.g. theoretically advantaged but perform poorly) but not unrealistic. Therefore, the results from their removal seem to reflect the consequence of removing heterogeneity and important subgroups within the sample, rather than outlier data likely to be based on errors.

Table 3.30. Extreme outliers descriptive statistics

	Box plot outliers						Cook's distance outliers			
	BAS-2 NV scores of 79 or 80 (N = 59)		FSPT scores of less than 40 (N = 69)		Both BAS-2 NV and FSP total outliers (N = 128)		Model 4 (N = 338)		Model 5 (N = 347)	
Measure	**% (N)**	**Mean (SD)**	**% (N)**	**Mean (SD)**	**% (N)**	**Mean (SD)**	**% (N)**	**Mean (SD)**	**% (N)**	**Mean (SD)**
BAS-2 NV	-	79.69 (0.50)	-	38.38 (9.00)	-	57.38 (21.66)	-	43.18 (11.62)	-	43.62 (12.16)
FSP total	-	100.20 (10.88)	-	32.90 (7.10)	-	63.92 (34.86)	-	66.38 (29.89)	-	66.21 (29.53)
Gender assigned at birth		-		-		-		-		-
Male	47.46 (28)		75.36 (52)		62.50 (80)		58.88 (199)		58.21 (202)	
Female	52.54 (31)		24.64 (17)		37.50 (48)		41.12 (139)		41.79 (145)	
Relative poverty (OECD 60% poverty threshold)		-		-		-		-		-
Below threshold	15.25 (9)		60.87 (42)		39.84 (51)		52.07 (176)		54.18 (188)	
Above Threshold	84.75 (50)		39.13 (27)		60.16 (77)		47.93 (162)		41.79 (159)	

Age (in months)	-	38.11 (2.57)	-	38.31 (2.80)	-	38.22 (2.69)	-	38.56 (3.44)	-	38.66 (3.48)
Initial school readiness (BSRA-R)	-	124.92 (11.07)	-	85.55 (13.66)	-	103.70 (23.32)	-	94.91 (16.76)	-	95.12 (16.89)
Behaviour (SDQ total difficulties)	-	6.36 (3.95)	-	13.38 (6.13)	-	10.14 (6.29)	-	11.71 (6.01)	-	11.74 (6.02)
Maternal education (NVQ level)										
No qualifications	3.39 (2)		26.09 (18)		15.63 (20)		23.96 (81)		24.50 (85)	
NVQ 1	3.39 (2)		24.64 (17)		14.84 (19)		16.27 (55)		16.14 (56)	
NVQ 2	22.03 (13)		33.33 (23)		28.13 (36)		26.04 (88)		25.94 (90)	
NVQ 3	13.56 (8)		5.80 (4)		9.38 (12)		10.36 (35)		10.66 (37)	
NVQ 4	44.07 (26)		10.14 (7)		25.78 (33)		17.75 (60)		17.00 (59)	
NVQ 5	13.56 (8)		0 (0)		6.25 (8)		5.62 (19)		5.76 (20)	
Maternal age at birth (years and months)	-	30.66 (4.94)	-	26.17 (6.24)	-	28.24 (6.09)	-	27.81 (6.61)	-	27.70 (6.34)

Maternal mental health (Kessler-6)	-	2.59 (2.46)	-	5.15 (5.77)	-	3.97 (4.72)	-	4.65 (5.06)	-	4.72 (5.12)
Home learning environment (HLE index)	-	30 (6.66)	-	22.90 (7.91)	-	26.17 (8.15)	-	24.65 (7.86)	-	24.96 (7.80)

Note. %(N) are to outline the proportions for specific measures, total N for identified outliers are presented under each outlier type

Chapter 4. Discussion: The influence of child and social factors on the efficacy of language interventions and their role as moderators of the effect of language on school readiness

4.1. Research issue and questions

A large volume of literature indicates that preschool oral language and school readiness are key developmental attainments, and important indicators for many longer-term life and societal outcomes (Bishop, 2009; Bishop et al., 2016; Botting et al., 2016; Curtis et al., 2018, 2019; S. Davies et al., 2016; Feeney et al., 2012; Johnson et al., 2010; Jones et al., 2015; Klem et al., 2016; Law, 2015; McKean et al., 2017; Pan et al., 2019; Rahman et al., 2018; Reynolds et al., 2011; Ricciardi et al., 2021; Sadler et al., 2015; van den Bedem et al., 2018). Language ability is also a subcomponent of school readiness, and is associated with the other skills that are subcomponents of school readiness (Bretherton et al., 2014; Chow et al., 2018; Chow & Ekholm, 2019; Fuchs et al., 2018; Levickis et al., 2018; Lonigan & Milburn, 2017; Snijders et al., 2020; Trakulphadetkrai et al., 2020; Willinger et al., 2019; Yew & O'Kearney, 2013). (Bretherton et al., 2014; Chow et al., 2018; Chow & Ekholm, 2019; Fuchs et al., 2018; Levickis et al., 2018; Lonigan & Milburn, 2017; Snijders et al., 2020; Trakulphadetkrai et al., 2020; Willinger et al., 2019; Yew & O'Kearney, 2013). Due to this, efforts to boost school readiness through early language is advocated for, and implemented successfully (EEF, 2019; Law et al., 2018; Lonigan et al., 2015; Nix et al., 2013; Noble et al., 2012; Welsh et al., 2020). But while intervention research demonstrates that preschool language interventions can benefit oral language and school readiness, their implementation and examination of effects tends to be motivated by set of implicit assumptions. Specifically, they assume 1) children benefit equally from language interventions (examined in chapter 2), and 2) children will benefit equally in school readiness outcomes form gains in language ability (examined in chapter 3).

However, research has indicated specific developmental vulnerabilities and social characteristics place children at risk of poor school readiness and language outcomes (Betancourt et al., 2015; Duncan et al., 2015; Duncan et al., 2007; Flouri et al., 2020; Hosokawa & Katsura, 2018; Levickis et al., 2018; Neuman et al., 2018;

Paul, 2020; Vugteveen et al., 2021). As such, there may be numerous possible pathways or mechanisms through which oral language, school readiness, child and social factors associate with one-another. Some reviews and intervention studies have indicated that intervention response may differ due to certain developmental vulnerabilities and social disadvantages (Boyle et al., 2007; Dowdall et al., 2020; Marulis & Neuman, 2010, 2013; Roberts & Kaiser, 2011; Storkel et al., 2017). Research also indicates that not only do risk factors have an effect on school readiness independent of language, but they may also affect the ability of children to capitalise on initial language advantages (Feinstein, 2003; Hammer et al., 2017; Prior et al., 2011). When taken together, a further concern is that children may be subject to a 'triple threat' of disadvantage. That is, oral language and school readiness may be affected by 1) direct effects of social disadvantage and developmental vulnerabilities, 2) poorer response to language interventions and 3) less benefit accrued for school readiness from language gains. In other words, children with developmental vulnerabilities and social disadvantages could be receiving a cumulative disadvantage towards their language and school readiness development, gains from intervention, and gains in school readiness even if they benefit from interventions. Therefore, employing current interventions without considering how to tackle these levels of disadvantage may compound difficulties that some children already demonstrate.

However, the current evidence available to test these hypotheses is limited and a number of research gaps were identified. For language intervention efficacy, the pool of child and social factors examined were limited in studies, and findings for most factors were from small samples, and/or a small number of studies. The effects of child and social factors on intervention efficacy were also generally not the focus of the studies. In addition, no research to my knowledge examines the potential moderating effect of child and social factors on the relationship between oral language and school readiness. Furthermore, no study to my knowledge has explored the 'triple threat' of disadvantage outlined here. Therefore, there was a need to complete comprehensive research which focuses specifically on the effects of child and social factors on intervention response, and how for child and social factors moderate the relationship between language and school readiness. This was done by answering an overarching question: "*To what extent do child and social factors* *moderate the efficacy of language interventions, and what is their role as moderators of the effect of language on school readiness?",* split into the two following questions:

1) Do children benefit equally from interventions, or are gains affected by child and social factors?
2) Do children benefit equally in their school readiness outcomes from gains in language ability, or are these benefits moderated by additional child and social factors?

4.2. Key findings

4.2.1. Phase 1: Do children benefit equally from interventions or are gains affected by child and social factors?

Findings indicated that differences in initial language and speech skills affected differential intervention response. In addition, area deprivation/free school meals, NVIQ, and age were shown to affect language growth.

Intervention response was better for children with more severe initial language difficulties for general language, word knowledge, and expressive morphosyntax. These findings ran counter to the theory that children would gain less from interventions as their weaker language skills would make it harder for them to understand and engage in some steps/tasks related to those skills in the intervention. This may in turn prevent them from gaining the maximum possible benefit from what is being taught (Storkel et al., 2017). Instead, it is clear that interventions are effectively targeting gaps for children with more severe difficulties, and it may be that children with milder difficulties gain less because they have less gains to make. Nevertheless, children with milder difficulties were still shown to benefit from interventions. This is good news because while these different subgroups show gains, each examined subgroup is still benefitting from intervention. In comparison to the other oral language skills, children with milder language difficulties gained more from interventions than children with more severe difficulties in a study using covariate analysis. As discussed in chapter 2, the difference seen with initial listening comprehension skills for the covariate analyses may be due to listening comprehension being supported by cognitive processing skills and syntactic and vocabulary skills (Kim & Pilcher, 2016). As such, listening comprehension may be harder to treat for children with more severe listening comprehension difficulties as they may also have more general language difficulties. Finally, there were non-significant moderation findings for initial language for expressive vocabulary, receptive vocabulary, mixed morphosyntax and semantics, and phonological awareness outcomes. This is potentially good news, as children benefit equally in interventions addressing these outcomes regardless of their level of difficulty.

However, as noted in chapter 2 it could be that some differences found (specifically for word knowledge and expressive morphosyntax) are not true effects due to not being RCTs, and instead may be due to regression to the mean (Linden, 2013). In addition, as discussed in the strengths and limitations section, the availability and quality of the evidence for the effects of initial language was poor. Therefore, caution should be applied when drawing conclusions from these results. However, there appears to be a clear pattern that the severity of a child's language difficulty may shape their gains from language intervention. The direction and effect may also differ depending on the initial language skill(s) the child has difficulty in, and the language skill outcome. As such, it would be worth further examining the effects of language severity on intervention response, and for different language skills in future research.

Speech was found to predict outcome growth and create differential intervention response for phonological awareness and expressive morphosyntax. Specifically, children with better speech skills benefitted more from language interventions, and those with worse speech benefitted less. This supported the hypothesis posited that because weaker speech undermines oral language development (Haskill & Tyler, 2007; Lewis et al., 2015), this could also potentially undermine the extent of children's gains in their language via intervention. Nevertheless, where effect sizes for this subgroup could be obtained (1 study), children with poor speech still benefitted moderately. Like for the findings with initial language, this is encouraging because while children demonstrate differential benefits based on their speech, they are all still benefitting from intervention. As mentioned in chapter 2, speech was only examined in interventions with phonological awareness and expressive morphosyntax. Both of these language skills are strongly related to speech (Dodd et al., 2018; Murray et al., 2019). As such, more research needs to be completed to determine if these findings are indicating an actual differential response, or based on incorrectly conflating the child's language knowledge with their ability to signal phonemes or morphemes expressively.

Three other child and social factors were identified to potentially influence language growth during intervention. Area deprivation/free school meal uptake introduced 'noise' to gains in mixed morphosyntax and semantics and phonological awareness, but its effects were unclear. As such, it could not be determined if findings supported the hypothesis made that because language interventions can address the deficits in resources promoting language at home, children living in social disadvantage could benefit more from language interventions (McKean et al., 2015, 2017). However, area deprivation/free school meals appears to have some type of influence on language growth in the intervention. Therefore, it would be worth further examining the effects of this factor on intervention response in future research.

In line with prior literature (Ebert, 2021; Griffiths et al., 2022; Smolak et al., 2020; Snijders et al., 2020; Willinger et al., 2019; Yim & Yang, 2018), better non-verbal IQ appeared to positively predict general language, expressive and receptive vocabulary gains from interventions. However, NVIQ was a non-significant moderator for expressive vocabulary and word knowledge intervention gains. This finding is counter to the hypothesis made that children scoring lower on NVIQ assessments have more general cognitive difficulties which provide a barrier to their engagement with learning activities (Alibali & Nathan, 2018), and weakens their language development (Griffiths et al., 2022) and resultant ability to engage with language-based learning; therefore gaining less because they find it difficult to engage with learning tasks in interventions. Instead, findings support the previous research seen for older children, that NVIQ does not impact language intervention response (Boyle et al., 2007). Why this finding occurred could be because the study using moderation analyses ensured their tasks were suitable to children with different cognitive profiles. Therefore, this could have removed the hypothesised barrier of engaging with learning activities that children scoring lower on NVIQ assessments may have. However, moderation analyses were conducted for one study, and for word knowledge and expressive vocabulary gains only. As such, it would be worth further examining the effects of this factor on intervention response in future research for different language outcomes.

Being older was predictive of better growth in an intervention for phonological awareness. This supports the hypothesis that older children may benefit more in interventions because they are more experienced in educational tasks; and they are generally more cognitively and socially developed which allows them to access learning more easily (Cantalini-Williams et al., 2016). However, other studies examining age did not find significant differences between subgroups for general language and expressive and receptive vocabulary. Why age was significant only for phonological awareness may be because phonological awareness is a metalinguistic skill, and may require developmental maturity in order to access further learning for this skill (Gombert, 1997). This also supports research that children at or older than 5 may still be able to benefit significantly from language intervention (McKean et al., 2015; Taylor et al., 2013).

Finally, language profile, behaviour, maternal education, gender assigned at birth and non-specific difficulties were not clearly or significantly related to intervention response or outcome growth. This is potentially encouraging as it could indicate that children may benefit equally in interventions regardless of their differences in these factors. However, the intervention effects were non-significant in studies these were included in, and findings are based on a small number of studies (usually 1) and very low quality of evidence. As such, it cannot be ruled out that these child and social factors do potentially have an effect on intervention response.

4.2.2. Phase 2: Do children benefit equally in their school readiness outcomes from gains in language ability or are these benefits moderated by additional child and social factors?

Better expressive vocabulary predicted a better total FSP score and better scores for each sub skill measured by the scale (personal, social and emotional development; communication, language and literacy; mathematical development; knowledge and understanding of the world; creative development; physical development) as part of the school readiness construct. Children with better vocabulary were also more likely to achieve school readiness based on a government benchmark (Good Level of Development). This was expected according to prior literature (George et al., 2007; Rodriguez & Tamis-LeMonda, 2011), indicating that having good language predicts children being more ‘school ready’.

Gender assigned at birth and poverty were chosen as moderators for the longitudinal analysis due to their theoretical fit (their relation to language development and robust and direct influence on school readiness) and measurement quality. For gender assigned at birth, males gained more in school readiness if they had good language. For poverty, children living in poverty gained more in school readiness if they had good language. However, the differences between males and females, and those living above and below the poverty threshold were small. While good language appears to be a protective factor for these at-risk groups, children living in poverty and assigned male at birth may benefit more as they use this to compensate for their developmental and social disadvantages; while females can draw on their developmental advantages, and more affluent children can draw on their resource advantages to access learning relatively well (gender assigned at birth: Adani & Cepanec, 2019; Talbot, 2020, children in poverty: Duncan et al., 2014; Hobcraft & Kiernan, 2010; Illøkken et al., 2021; Mollborn et al., 2014). Thus, language gains still have an effect on school readiness gains for both females and more affluent children, but to a slightly lesser degree. Therefore, it is essential children in at risk groups receive language intervention to ensure good oral language so they can capitalize on any skill which can give them similar outcomes to others. Children achieved similarly in the government benchmark of school readiness (FSP GLD) in gender assigned at birth or poverty subgroups. It is unclear why there is a difference between this and the other two measurements, but it is likely converting achievement to a binary variable is reductive as it decreases sensitivity to the variety of individual differences children have when beginning school. Therefore, it is likely it may also not be sensitive to whether school readiness benefits from gains in language are moderated by additional child and social factors.

Gender assigned at birth did not predict or moderate effects for the knowledge and understanding of the world and physical development school readiness concepts. Poverty did not predict or moderate effects for the personal, social and emotional development, knowledge and understanding of the world, and creative development school readiness concepts. Due to the learning tasks which make up creative development and knowledge and understanding of the world, they may be less reliant on developmental and/or resource advantages. For physical development on the other hand, some of the early learning goals reflect domains that females have alternative developmental advantages in, and so this may make up for their theorised developmental disadvantage in physical skills. For personal, social and emotional development, the effects of having stronger language may be an important protective factor for socio-emotional risks seen in children living in poverty. Therefore, engaging children’s interests and promoting their language may drive their ‘readiness’ in these skills rather than resources or developmental advantages (directly related to the skill). On the other hand, it may also be worth examining closely what skills early learning goals are assessing, and whether these could be helping children compensate for other developmental disadvantages that may be important to take note of.

4.2.3. The ‘triple threat’ of disadvantage

From the evidence collected, it is difficult to verify whether the child and social factors explored in this thesis would present children with a ‘triple threat’ of disadvantage for all factors examined. This is because factors identified in the first phase could not be utilized as moderators in the second phase due to their quality or availability as a measure in the longitudinal dataset (e.g., speech difficulties). Even if available, oral language skills examined were usually different (e.g., speech was examined with reference to phonological awareness and expressive morphosyntax outcomes, the second phase utilized expressive vocabulary as the predictor), so there could not be a 1:1 mapping of oral language skills for most factors (e.g., speech difficulties directly affect expressive vocabulary development, impact expressive vocabulary gains from intervention, and impact school readiness gains from having good expressive vocabulary).

However, conclusions could be tentatively drawn for gender assigned at birth. Moderation analyses indicated that there was no difference for intervention response for expressive vocabulary. Furthermore, children benefitted differently from having good expressive vocabulary for school readiness, but benefits were in favour of children in the ‘at risk’ group (males). Together, this indicates that being male does not present children with a ‘triple threat’ of disadvantage. In addition, findings indicated speech difficulties could create at least a ’double threat’ of disadvantage (i.e., speech difficulties directly affect phonological awareness and expressive morphosyntax development, and impact phonological awareness and expressive morphosyntax gains from intervention).

4.3. Strengths and limitations

The systematic review to the author's knowledge is the first systematic examination of the current available evidence of analyses exploring how child and social factors may produce differential intervention response. It was also able to highlight research gaps in conducting and reporting such analyses. Furthermore, the review included a number of different child and social factors, and considered a range of analytical approaches studies utilised by researchers for child and social factors. Limitations from the systematic review concerned the availability and quality of the evidence. Findings for each factor and language outcome were based on a small number of studies, and significant results came from a very small number of mostly quasi-experimental studies. The quality of the studies and hence the confidence that can be had in the findings was also low or very low in most aspects (risk of bias in studies, inconsistency, publication bias and imprecision). Effect sizes for analyses relating to differential intervention response based on child and social factors were not possible to calculate for most studies. As such, evidence from the systematic review is very tentative and should be interpreted and applied with caution as there is an absence of evidence that has been sufficiently robustly tested. However, the findings do provide a springboard for future research, by 1) providing a set of hypotheses for factors which researchers can begin to expand upon, 2) recommending how to report such analyses, and 3) recommending what data to present in intervention studies to allow for meta-analyses (see section 4.4).

The secondary data analysis to the author's knowledge is the first to examine how child and social factors moderate the effects of expressive vocabulary on school readiness. Its findings reflect a robust analysis of a nationally representative longitudinal cohort study with over 5,500 children. The highest quality variables available for factors of interest were selected based on a thorough examination process. However, the dataset had some limitations. Only expressive vocabulary data was collected, meaning findings were limited in their application to other oral language skills. However, vocabulary is a good indicator of broader preschool language development (Bishop et al., 2017; Tomblin & Zhang, 2006) and so findings likely reflect broader child language ability. Furthermore, previous longitudinal research has yielded significant insights into the long-term impacts of vocabulary on language and skills included as part of the school readiness construct (Coloma et al., 2020; Westrupp et al., 2020; Willoughby, 2020). Certain exclusions based on data collected in expressive vocabulary, maternal education and initial school readiness meant children who were multilingual or had special educational needs were likely under-represented. As such, data which includes these children needs to be obtained in future longitudinal research with measures also suited to them. Furthermore, while multiple imputation might have offered some advantages, this was not possible within the time and resource constraints of this thesis. This meant that children were left out of the final moderation analyses, and these children were more likely to be from socially disadvantaged groups. While not fully ameliorating the underestimation of social poverty effects, applying the sample weights of the MCS was able to allow analyses to be more representative and account for missing data.

4.4. Recommendations

4.4.1. Research

This thesis has highlighted the need for more high-quality research examining child and social factors as moderators in language intervention; and to determine how they moderate the benefits of school readiness from language gains. There are three ways this question can be addressed. The first is to conduct language intervention efficacy studies which examine this specifically, the second is exploring moderation effects via meta-analyses, and the third is examining predictive interactive models in longitudinal datasets.

For language intervention efficacy studies, these need to explicitly describe and assess the effects of additional child and social characteristics. In order for robust conclusions to be drawn, researchers need to select theoretically-based characteristics *a priori*, and utilise high quality measures for chosen factors. This will ensure that researchers are actively considering which child and social characteristics are making gains from interventions stronger or weaker. In addition, it will mean that conclusions drawn can be done so with confidence. Furthermore, research needs to be utilising predictive interactive models with high statistical power to ensure that evidence is valid and robust. To do so, research should utilise Randomised Control Trials so that conclusions can be based on individual growth and findings are less likely to be subject to regression to the mean.

Second, in order to conduct high-quality meta-analyses where data can be synthesized well, researchers should work to collaborate and create agreed upon guidelines for how to extract and synthesise data relating to social and child factors. This will then allow for high-quality, standardised evidence to be available, so recommendations made from findings can be in turn implemented with confidence.

Third, while the thesis has begun the necessary work to understand how child and social factors may moderate gains in school readiness from gains in language, more research is needed. Specifically, examining more child and social factors (including those highlighted throughout the thesis) and using high-quality and robust measures is needed. Moreover, measurement should also be inclusive of specific social groups (e.g., disabled children, children from multilingual backgrounds, etc.). In addition, it would also be important to determine if the effects found for poverty and gender assigned at birth can be replicated with other longitudinal datasets. Currently, there are many modern or ongoing developmental cohorts, with some examples being:

- The Early Language in Victoria study (Australia): Which includes around 1,900 families, and has examined language development in children from infancy to adolescence (Reilly et al., 2018);
- The Avon Longitudinal Study of Parents and Children (England): Which includes nearly 14,000 mothers and children, and has followed their health and developmental outcomes for nearly two decades. (Fraser et al., 2013);
- Growing up in Scotland (Scotland): Which includes around 3,000 families, and examines how social inequalities from infancy can impact later health and developmental outcomes for children and adolescents (CLOSER, 2022);
- The Quebec Longitudinal Study of Child Development (Canada): Which includes over 2,000 children, and followed their physical, cognitive, social, and emotional development from infancy to adulthood (Orri et al., 2021);
- The Early Childhood Longitudinal Study (US): Which included over 18,000 children, and followed their educational and socio-emotional outcomes from kindergarten to late childhood (NCES, 2022)
- The Longitudinal Study of Australian Children (Australia): Which includes around 10,000 children and their families, and has followed

their development and well-being from infancy to adulthood (Growing up in Australia, 2022).

Each of these datasets include a variety of language, educational, child, and social measures. As such, they could be useful to utilise in future research to understand how child and social factors may moderate gains in school readiness from gains in language. However, as can be seen, these datasets typically represent one country or union. As such, for findings to be generalizable internationally, there is a need for more multinational large-scale datasets that include a variety of language, school readiness, and child and social factors. Having this higher level of generalization will help feed into creating theoretically robust interventions, as they will be able to see how child and social factors moderate intervention response, and the benefits of school readiness from gains in language not just nationally, but internationally. This will help inform researchers, practitioners and policy makers how to best provide intervention for children with a variety of developmental vulnerabilities and social disadvantages.

If researchers are able to implement these recommendations for language interventions, longitudinal research, and meta-analyses, then this will also help determine how factors may relate in deeper ways (i.e., ‘triple threat’ of disadvantage) to affect children’s school readiness gains and intervention response.

4.4.2. Policy and practice

The findings from the systematic review and longitudinal analysis make it evident that employing language interventions is likely to be worthwhile for school readiness outcomes. Furthermore, assumptions should not be made by those overseeing or implementing interventions as to who benefits from them. Prior research highlighted throughout has suggested that children’s developmental vulnerabilities (being male, more severe language difficulties) and social disadvantages (living in poverty) may individually predict poorer language and school readiness outcomes. However, the findings of this thesis demonstrate children in certain risk groups could benefit more in language interventions; and in their school readiness from language gains compared to more advantaged peers. Therefore, employing interventions promoting language for at risk children could in turn improve their ability to be school ready. In addition, children with social or developmental advantages either still benefitted from language interventions; or still rely on their language gains to increase gains in school readiness, even if to a lesser extent. As such, if more advantaged children had language difficulties, they would also likely benefit from interventions promoting their language. However, children with developmental vulnerabilities and social disadvantages may rely more on intervention and good language to compensate for their social and developmental disadvantages. Therefore, policy prioritising language interventions for these children may be especially key to them achieving developmental milestones in language and school readiness. The longitudinal analysis findings also indicated certain components of school readiness may be more susceptible to the developmental and social differences between children than others. As such, educational policy should focus on assessing early learning goals and creating related tasks in the classroom which take into account the developmental and social disadvantages children have. This may in turn allow children from different developmental and social subgroups to access learning and achieve equitably in these aspects.

The systematic review findings indicated that children with both language and speech difficulties gained less from language interventions. It is therefore likely essential for children with language difficulties to also be assessed for speech difficulties and vice versa. This is so language interventions can also address speech difficulties where they occur, which in turn will likely improve response for these children. Furthermore, it appears that although receiving less benefit, children with speech difficulties still benefitted to a moderate extent. As such, it is clearly important to not exclude children with speech difficulties as language interventions are likely to be still beneficial to them. When considering age, there was little data to determine whether early or later intervention should be prioritised. However, the systematic review found that children entering school and older could also benefit from continued phonological awareness interventions. As such, it may also be important to bear in mind that intervention may not always be most optimal when targeted at very young children; and instead it may be better to employ phonological awareness interventions for older as well as younger children.

There is also a clear need to fund more large-scale longitudinal national and international cohort studies which includes child language development, and RCTs examining how child and social factors impact language intervention efficacy. These will provide more robust findings with which to inform policy and practice. Due to the costs of running these projects, there may be some hesitancy to fund multiple large-scale designs. As such, any funding awarded should allow researchers to conduct pilot studies (using quasi-experimental or case study designs). This is so researchers can obtain initial findings that are promising to replicate, and make changes if needed before the larger-scale research commences. Funding bodies should also ensure that they consult research experts and practitioners in the field to develop criteria to guarantee that the most realistic, practical and best quality research is funded. This would then maintain value for money and ensure funded studies will elevate the current knowledge in the field.

4.5. Contributions made by the research and conclusions

To my knowledge, this thesis completed the first comprehensive and focused investigation into how different developmental vulnerabilities and social disadvantages moderated 1) language intervention response, and 2) the benefits on school readiness from gains in expressive vocabulary. This thesis was conducted with the aim to assist researchers, practitioners and policy makers to have a deeper understanding of how the developmental and social inequalities that children may face impact their ability to respond to intervention, and capitalise on gains made to benefit their school readiness. The findings were the first to my knowledge which support the idea that language, school readiness, child, and social factors associate with one-another through complex mechanisms which may not just predict additive risk to language development and school readiness. Instead, these mechanisms may also operate through other interactive relationships. For example, the thesis was able to start to explore whether some risk factors created ‘triple’ or ‘double’ threats to children’s’ oral language and school readiness; and found when examining these relationships that they may not go in the way that is expected. Therefore, it is clear that there may be more complex relationships between child and social factors, language intervention and school readiness gains, and how these play out cannot be assumed based on additive models. More of these relationships need to be explored in future research, so we can understand how to ensure all children who require it can obtain equitable support in promoting their oral language and school readiness.

References

Adani, S., & Cepanec, M. (2019). Sex differences in early communication development: Behavioral and neurobiological indicators of more vulnerable communication system development in boys. *Croatian Medical Journal*, *60*(2). https://doi.org/10.3325/cmj.2019.60.141

Aiman, S., Yusof, S. M., Kadir, Z. A., & Sabturani, N. (2016). The Relationship Between Socioeconomic Status and Fine Motor Skills Among Six-Year-Old Preschool Children. In *Proceedings of the 2nd International Colloquium on Sports Science, Exercise, Engineering and Technology 2015 (ICoSSEET 2015)*. https://doi.org/10.1007/978-981-287-691-1_15

Akhter, S., Pauyo, T., & Khan, M. (2019). What Is the Difference Between a Systematic Review and a Meta-analysis? In *Basic Methods Handbook for Clinical Orthopaedic Research*. https://doi.org/10.1007/978-3-662-58254-1_37

Alibali, M. W., & Nathan, M. J. (2018). Embodied Cognition in Learning and Teaching: Action, observation and imagination. In F. Fischer, C. . Hmelo-Silver, S. . Goldman, & P. Reimann (Eds.), *International Handbook of Learning Sciences* (pp. 260–271). Routledge.

Ambridge, B., & Lieven, E. V. . (2011). *Child Language Acquisition: Contrasting Theoretical Concepts*. Cambridge University Press.

American Psychological Association. (2021). *APA style*. American Psychological Association. https://apastyle.apa.org/

Amorsen, A., & Miller, M. (2017). Children’s oral language development and early literacy practices. *Educating Young Children: Learning and Teaching in the Early Childhood Years*, *23*(1), 24–27.

Armon-Lotem, S., de Jong, J., & Meir, N. (2015). *Assessing multilingual children: Disentangling bilingualism for language impairment*. United Kingdom: Channel View Publications.

Assmusen, K., Fischer, F., Drayton, E., & McBride, T. (2020). *Adverse childhood experiences: What we know, what we don’t know, and what should happen next*. Early Intervention Foundation. https://www.eif.org.uk/report/adverse-childhood-experiences-what-we-know-what-we-dont-know-and-what-should-happen-next

Austin, A. (2018). Association of Adverse Childhood Experiences with Life Course Health and Development. In *North Carolina medical journal* (Vol. 79, Issue 2). https://doi.org/10.18043/ncm.79.2.99

Austin, P. C., & Stuart, E. A. (2017). Estimating the effect of treatment on binary outcomes using full matching on the propensity score. *Statistical Methods in Medical Research*, *26*(6). https://doi.org/10.1177/0962280215601134

Bakker, A., Cai, J., English, L., Kaiser, G., Mesa, V., & Van Dooren, W. (2019). Beyond small, medium, or large: points of consideration when interpreting effect sizes. In *Educational Studies in Mathematics* (Vol. 102, Issue 1). https://doi.org/10.1007/s10649-019-09908-4

Banerjee, A., Pluddemann, A., O'Sullivan, J., & Nunan, D. (2019). *Performance bias: Systematic differences in the care provided to members of different study groups other than the intervention under investigation*. Catalogue of Bias. https://catalogofbias.org/biases/performance-bias/

Bärnighausen, T., Tugwell, P., Røttingen, J. A., Shemilt, I., Rockers, P., Geldsetzer, P., Lavis, J., Grimshaw, J., Daniels, K., Brown, A., Bor, J., Tanner, J., Rashidian, A., Barreto, M., Vollmer, S., & Atun, R. (2017). Quasi-experimental study designs series—paper 4: uses and value. *Journal of Clinical Epidemiology*, *89*. https://doi.org/10.1016/j.jclinepi.2017.03.012

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6). https://doi.org/10.1037//0022-3514.51.6.1173

Bassett, H. H., Denham, S., Mincic, M., & Graling, K. (2012). The Structure of Preschoolers' Emotion Knowledge: Model Equivalence and Validity Using a Structural Equation Modeling Approach. *Early Education and Development*, *23*(3). https://doi.org/10.1080/10409289.2012.630825

Beauregard, J. L., Drews-Botsch, C., Sales, J. M., Flanders, W. D., & Kramer, M. R. (2018). Preterm birth, poverty, and cognitive development. *Pediatrics*, *141*(1). https://doi.org/10.1542/peds.2017-0509

Bernal, R., & Keane, M. P. (2011). Child care choices and children's cognitive achievement: The case of single mothers. *Journal of Labor Economics*, *29*(3).

https://doi.org/10.1086/659343

Betancourt, L. M., Brodsky, N. L., & Hurt, H. (2015). Socioeconomic (SES) differences in language are evident in female infants at 7 months of age. *Early Human Development*, *91*(12). https://doi.org/10.1016/j.earlhumdev.2015.08.002

Bishop, D. (2009). Specific language impairment as a language learning disability. *Child Language Teaching and Therapy*, *25*(2), 163–165. https://doi.org/10.1177/0265659009105889

Bishop, D. V. M. (2014). Ten questions about terminology for children with unexplained language problems. In *International Journal of Language and Communication Disorders* (Vol. 49, Issue 4). https://doi.org/10.1111/1460-6984.12101

Bishop, D. V, Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2016). CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PLOS ONE*, *11*(7), 1–26. https://doi.org/https://doi.org/10.1371/journal.pone.0158753

Bishop, D. V, Snowling, M. J., Thompson, P. A., & 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*(10), 1068–1080. https://doi.org/10.7287/PEERJ.PREPRINTS.2484V1

Bjerke, S. M., Feragen, K. B., & Bergvik, S. (2018). Strengths and difficulties questionnaire (SDQ): Informant agreement between children born with cleft lip and/or palate and their parents. *Cleft Palate-Craniofacial Journal*, *55*(2). https://doi.org/10.1177/1055665617730365

Blanden, J., & Machin, S. (2010). Intergenerational inequality in early years assessments. In *Children of the 21st Century: The First Five Years* (Vol. 2). https://doi.org/10.2307/j.ctt9qgrfx.17

BMJ publishing group. (2017). *What is GRADE?* BMJ Clinical Evidence.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2). https://doi.org/10.1002/jrsm.12

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). When Does it Make Sense to Perform a Meta-Analysis? In *Introduction to Meta-Analysis*. https://doi.org/10.1002/9781119558378.ch45

Bothwell, L. E., Greene, J. A., Podolsky, S. H., & Jones, D. S. (2016). Assessing the Gold Standard — Lessons from the History of RCTs. *New England Journal of Medicine*, *374*(22). https://doi.org/10.1056/nejmms1604593

Botting, N., Toseeb, U., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2016). Depression and anxiety change from adolescence to adulthood in individuals with and without language impairment. *PLoS ONE*, *11*(7), e0156678. https://doi.org/10.1371/journal.pone.0156678

Bouchard, C., Trudeau, N., Sutton, A., Boudreault, M. C., & Deneault, J. (2009). Gender differences in language development in French Canadian children between 8 and 30 months of age. *Applied Psycholinguistics*, *30*(4). https://doi.org/10.1017/S0142716409990075

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 Technology Assessment*, *11*(25). https://doi.org/10.3310/hta11250

Boyle, J., Mccartney, E., O'hare, A., & Law, J. (2010). Intervention for mixed receptive-expressive language impairment: A review. In *Developmental Medicine and Child Neurology* (Vol. 52, Issue 11). https://doi.org/10.1111/j.1469-8749.2010.03750.x

Bradshaw, J., & Holmes, J. (2010). Child poverty in the first five years of life. In *Children of the 21st Century: The First Five Years* (Vol. 2). https://doi.org/10.2307/j.ctt9qgrfx.10

Bretherton, L., Prior, M., Bavin, E., Cini, E., Eadie, P., & Reilly, S. (2014). Developing relationships between language and behaviour in preschool children from the Early Language in Victoria Study: implications for intervention. *Emotional and Behavioural Difficulties*, *19*(1). https://doi.org/10.1080/13632752.2013.854956

Brilli, Y., Boca, D. Del, & Monfardini, C. (2013). Child care arrangements: determinants and consequences. In *Families and Societies* (Vol. 2).

Brook, R. J., & Arnold, G. C. (2018). Applied Regression Analysis and Experimental Design. In *Applied Regression Analysis and Experimental Design*. https://doi.org/10.1201/9781315137674

Broomfield, J., & Dodd, B. (2011). Is speech and language therapy effective for children with primary speech and language impairment? Report of a randomized control trial. *International Journal of Language and Communication Disorders*, *46*(6). https://doi.org/10.1111/j.1460-6984.2011.00039.x

Bruce Tomblin, J., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research*, *49*(6). https://doi.org/10.1044/1092-4388(2006/086)

Bukodi, E., & Goldthorpe, J. H. (2013). Decomposing "social origins": The effects of parents' class, status, and education on the educational attainment of their children. *European Sociological Review*, *29*(5), 1024–1039. https://doi.org/10.1093/esr/jcs079

Burchinal, M., Magnuson, K., Powell, D., & Hong, S. S. (2015). Early childcare and education. In M. H. Bornstein, T. Leventhal, & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Ecological settings and processes* (pp. 223–267). John Wiley & Sons, Inc.

Bustamante, A. S., White, L. J., & Greenfield, D. B. (2017). Approaches to learning and school readiness in Head Start: Applications to preschool science. *Learning and Individual Differences*, *56*. https://doi.org/10.1016/j.lindif.2016.10.012

Bylund, E. (2009). Maturational constraints and first language attrition. *Language Learning*, *59*(3). https://doi.org/10.1111/j.1467-9922.2009.00521.x

Cabell, S. Q., Gerde, H. K., Hwang, H. J., Bowles, R., Skibbe, L., Piasta, S. B., & Justice, L. M. (2021). Rate of Growth of Preschool-Age Children's Oral Language and Decoding Skills Predicts Beginning Writing Ability. *Early Education and Development*. https://doi.org/10.1080/10409289.2021.1952390

Callahan, S., & Nicholas, L. (2019). Dragon wings and butterfly wings: implicit gender binarism in early childhood. *Gender and Education*, *31*(6). https://doi.org/10.1080/09540253.2018.1552361

Camacho, C., Straatmann, V. S., Day, J. C., & 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), 1–11. https://doi.org/10.1136/bmjopen-2018-024851

Cantalini-Williams, M., Perron, J., & Biemiller, A. (2016). Revisiting the Age-Old Question: What Are the Effects of Relative Age and Gender on Young Children in School Settings? *Journal of Childhood Studies*, *41*(2). https://doi.org/10.18357/jcs.v41i2.16095

Centre for Disease Control and Prevention. (2020). *Preventing Adverse Childhood Experiences*. https://www.cdc.gov/violenceprevention/childabuseandneglect/aces/fastfact.html

Centre for Longitudinal Studies. (2020). *Millennium Cohort Study: User Guide (Surveys 1-5)*. https://cls.ucl.ac.uk/wp-content/uploads/2020/09/MCS1-5_User_Guide_ed9_2020-08-07.pdf

Centre for Longitudinal Studies, & GfK NOP Social Research. (2006a). *MCS Age 3 Technical Report: Millennium Cohort Study*. Centre for Longitudinal Studies. https://cls.ucl.ac.uk/wp-content/uploads/2017/07/Millennium-Cohort-Study-Second-Survey-Technical-Report-on-Fieldwork-NOP-2006.pdf

Centre for Longitudinal Studies, & GfK NOP Social Research. (2006b). *MCS Age 3 Technical Report Appendicies: Millennium Cohort Study*. https://cls.ucl.ac.uk/wp-content/uploads/2017/07/Millennium-Cohort-Study-Second-Survey-Technical-Report-on-Fieldwork-NOP-Appendices-2006.pdf

Chen, E., & Miller, G. E. (2015). Early-life socioeconomic status, emotion regulation, and the biological mechanisms of disease across the lifespan. *Handbook of Emotion Regulation*.

Chow, J. C., & Ekholm, E. (2019). Language domains differentially predict mathematics performance in young children. *Early Childhood Research Quarterly*, *46*. https://doi.org/10.1016/j.ecresq.2018.02.011

Chow, J. C., Ekholm, E., & Coleman, H. (2018). Does oral language underpin the development of later behavior problems? A longitudinal meta-analysis. *School Psychology Quarterly*, *33*(3). https://doi.org/10.1037/spq0000255

Clifford, B. N., Stockdale, L. A., Coyne, S. M., Rainey, V., & Benitez, V. L. (2021).

Speaking of State of Mind: Maternal Mental Health Predicts Children's Home Language Environment and Expressive Language. *Journal of Child Language*. https://doi.org/10.1017/S0305000921000131

CLOSER. (2022). *Growing Up in Scotland*. https://www.closer.ac.uk/study/growing-up-in-scotland/

Coe, R., Kime, S., Neville, C., & Coleman, R. (2013). *The DIY evaluation guide.* https://educationendowmentfoundation.org.uk/public/files/Evaluation_Guide/EEF_Evaluation_DIY_Evaluation_Guide.pdf

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1). https://doi.org/10.1037/0033-2909.112.1.155

Collett, C. (2017). Supporting communciation. In C. Collett (Ed.), *Disability and Inclusion in Early Years Education*. Routledge.

Coloma, C. J., De Barbieri, Z., Quezada, C., Bravo, C., Chaf, G., & Araya, C. (2020). The impact of vocabulary, grammar and decoding on reading comprehension among children with SLI: a longitudinal study. *Journal of Communication Disorders*, *86*. https://doi.org/10.1016/j.jcomdis.2020.106002

Connelley, R. (2013). *Millennium Cohort Study Data Note: Interpreting Test Scores*. https://cls.ucl.ac.uk/wp-content/uploads/2017/06/Data-Note-20131_MCS-Test-Scores_Roxanne-Connelly-revised.pdf

Crowe, K., Cuervo, S., Guiberson, M., & Washington, K. N. (2021). A Systematic Review of Interventions for Multilingual Preschoolers With Speech and Language Difficulties. *Journal of Speech, Language, and Hearing Research*, *64*(11). https://doi.org/10.1044/2021_jslhr-21-00073

Cullis, A., & Hansen, K. (2008). *Child development in the first three sweeps of the Millennium Cohort Study.*

Curtis, P. R., Frey, J. R., Watson, C. D., Hampton, L. H., & Roberts, M. Y. (2018). Language disorders and problem behaviors: A meta-analysis. *Pediatrics*, *142*(2), e20173551. https://doi.org/10.1542/peds.2017-3551

Curtis, P. R., Kaiser, A. P., Estabrook, R., & Roberts, M. Y. (2019). The Longitudinal Effects of Early Language Intervention on Children's Problem Behaviors. *Child Development*, *90*(2), 576–592. https://doi.org/10.1111/cdev.12942

Dahlberg, A., Ghaderi, A., Sarkadi, A., & Salari, R. (2019). SDQ in the Hands of Fathers and Preschool Teachers—Psychometric Properties in a Non-clinical Sample of 3–5-Year-Olds. *Child Psychiatry and Human Development*, *50*(1). https://doi.org/10.1007/s10578-018-0826-4

Daily, S., Burkhauser, M., & Halle, T. (2010). A Review of School Readiness Practices in the States: Early Learning Guidelines and Assessments. *Child Trends*, *1*(3).

Daniel, G. R., & McLeod, S. (2017). Children with speech sound disorders at school: Challenges for children, parents and teachers. *Australian Journal of Teacher Education*, *42*(2). https://doi.org/10.14221/ajte.2017v42n2.6

Davies, N. M., Dickson, M., Smith, G. D., Van Den Berg, G. J., & Windmeijer, F. (2018). The causal effects of education on health outcomes in the UK Biobank. *Nature Human Behaviour*, *2*(2). https://doi.org/10.1038/s41562-017-0279-y

Davies, S., Janus, M., Duku, E., & Gaskin, A. (2016). Using the Early Development Instrument to examine cognitive and non-cognitive school readiness and elementary student achievement. *Early Childhood Research Quarterly*, *35*. https://doi.org/10.1016/j.ecresq.2015.10.002

de La Rochebrochard, E. (2012). Millennium Cohort Study Data Note 1: The home learning environment as measured at age 3. In *Centre for Longitudinal Studies*. https://cls.ucl.ac.uk/wp-content/uploads/2017/07/FINAL-The-home-learning-environment-as-measured-in-the-MCS-at-age-3-MCS-data-note-1.pdf

Deák, G. O. (2014). Interrelationship of Language and Cognitive Development (Overview). In *Encyclopedia of Language Development*. https://doi.org/10.4135/9781483346441.n91

Del Boca, D., Monfardini, C., & See, S. (2018). Government Education Expenditures, Pre-Primary Education and School Performance : A Cross-Country Analysis. *IZA Discussion Paper*, *11375*.

Department for education. (2020). *Early years foundation stage profile: 2021 handbook*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/942421/EYFSP_Handbook_2021.pdf

Department for Education. (2010). *Early years foundation stage profile results in England, 2010/11*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/219023/main_20text_20sfr282011.pdf

Department for Education and Skills. (2003). *Foundation Stage Profile Handbook*. http://doc.ukdataservice.ac.uk/doc/6847/mrdoc/pdf/foundation_stage_profile_handbook.pdf

Department for Education and Skills. (2004). *Statistics of education: Education and training statistics for the United Kingdom*. https://dera.ioe.ac.uk//4399/1/v06-2004.pdf

Dethorne, L. S., & Watkins, R. V. (2006). Language abilities and nonverbal IQ in children with language impairment: Inconsistency across measures. *Clinical Linguistics and Phonetics*, *20*(9). https://doi.org/10.1080/02699200500074313

Dex, S. (2008). Millennium Cohort Study: Exploration of some Distinctive Results for Scotland. In *Scottish Government Report*.

Dickerson, A., & Popli, G. K. (2016). Persistent poverty and children’s cognitive development: Evidence from the UK Millennium Cohort Study. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *179*(2). https://doi.org/10.1111/rssa.12128

Dodd, B., Reilly, S., Ttofari Eecen, K., & Morgan, A. T. (2018). Articulation or phonology? Evidence from longitudinal error data. *Clinical Linguistics and Phonetics*, *32*(11). https://doi.org/10.1080/02699206.2018.1488994

Dowdall, N., Melendez-Torres, G. J., Murray, L., Gardner, F., Hartford, L., & Cooper, P. J. (2020). Shared Picture Book Reading Interventions for Child Language Development: A Systematic Review and Meta-Analysis. In *Child Development* (Vol. 91, Issue 2). https://doi.org/10.1111/cdev.13225

Duncan, G.J., Magnuson, K., & Votruba-Drzal, E. (2015). Children and socioeconomic status. . In *Handbook of child psychology and developmental science: Ecological Settings and Processes* (Vol. 4, pp. 534–573).

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

Duncan, G.J., Magnuson, K., & Votruba-Drzal, E. (2014). Boosting family income to promote child development. *Future of Children*, *24*(1). https://doi.org/10.1353/foc.2014.0008

Eadie, P., Levickis, P., McKean, C., Westrupp, E., Bavin, E. L., Ware, R. S., Gerner, B., & Reilly, S. (2022). Developing Preschool Language Surveillance Models - Cumulative and Clustering Patterns of Early Life Factors in the Early Language in Victoria Study Cohort. *Frontiers in Paediatrics*. https://www.frontiersin.org/articles/10.3389/fped.2022.826817/full

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 and Child Neurology*, *57*(6). https://doi.org/10.1111/dmcn.12635

Early Years Matters. (2008). *Statutory Framework for the Early Years Foundation Stage*. https://earlyyearsmatters.co.uk/wp-content/uploads/2011/01/Statutory-Framework-for-the-Early-Years-Foundation-Stage-EYM.pdf

Ebbels, S. H., Marič, N., Murphy, A., & Turner, G. (2014). Improving comprehension in adolescents with severe receptive language impairments: A randomized control trial of intervention for coordinating conjunctions. *International Journal of Language and Communication Disorders*, *49*(1). https://doi.org/10.1111/1460-6984.12047

Ebbels, S. H., Wright, L., Brockbank, S., Godfrey, C., Harris, C., Leniston, H., Neary, K., Nicoll, H., Nicoll, L., Scott, J., & Marić, N. (2017). Effectiveness of 1:1 speech and language therapy for older children with (developmental) language disorder. *International Journal of Language and Communication Disorders*, *52*(4). https://doi.org/10.1111/1460-6984.12297

Ebert, K. D. (2021). Revisiting the Influences of Bilingualism and Developmental Language Disorder on Children’s Nonverbal Processing Speed. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/2021_jslhr-21-00156

EEF. (2019). *Oral language interventions*. Education Endownment Foundation Website. https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/oral-language-interventions/

Elardo, R., & Bradley, R. H. (1981). The home observation for measurement of the environment (HOME) scale: A review of research. In *Developmental Review* (Vol. 1, Issue 2). https://doi.org/10.1016/0273-2297(81)90012-5

Elliott, C. D. (1996). *The British Ability Scales II.* NFER-NELSON Publishing Company .

Enderby, P., John, A., Fryer, K., Cantrell, A., & Papaioannou, D. (2013). *Resource Manual for Commissioning and Planning Services for Speech, Language and Communication Needs*. https://www.rcslt.org/-/media/Project/RCSLT/resource-manual-dementia.pdf

Engel de Abreu, P. M. J., Baldassi, M., Puglisi, M. L., & Befi-Lopes, D. M. (2013). Cross-linguistic and cross-cultural effects on verbal working memory and vocabulary: Testing language-minority children with an immigrant background. *Journal of Speech, Language, and Hearing Research*, *56*(2). https://doi.org/10.1044/1092-4388(2012/12-0079)

Ensminger, M., & Fotherill, K. (2003). A decade of measuring SES: What it tells us and where to go from here . In M. Bornstein & R. Bradley (Eds.), *Socioeconomic status, parenting, and child development.* (pp. 13–27). Psychology Press.

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., Marjanovič-Umek, L., Gayraud, F., Kovacevic, M., & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology*, *30*(2). https://doi.org/10.1111/j.2044-835X.2011.02042.x

Erola, J., Jalonen, S., & Lehti, H. (2016). Parental education, class and income over early life course and children's achievement. *Research in Social Stratification and Mobility*, *44*, 33–43. https://doi.org/10.1016/j.rssm.2016.01.003

Feeney, R., Desha, L., Ziviani, J., & Nicholson, J. M. (2012). Health-related quality-of-life of children with speech and language difficulties: A review of the literature. *International Journal of Speech-Language Pathology*, *14*(1), 59–72. https://doi.org/10.3109/17549507.2011.604791

Feinstein, L. (2003). Inequality in the early cognitive development of British children in the 1970 cohort. *Economica*, *70*(277). https://doi.org/10.1111/1468-0335.t01-1-00272

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2). https://doi.org/10.1111/desc.12019

Fitzpatrick, C., & Pagani, L. S. (2012). Toddler working memory skills predict kindergarten school readiness. *Intelligence*, *40*(2). https://doi.org/10.1016/j.intell.2011.11.007

Flouri, E., Mavroveli, S., & Tzavidis, N. (2012). Cognitive ability, neighborhood deprivation, and young children’s emotional and behavioral problems. *Social Psychiatry and Psychiatric Epidemiology*, *47*(6). https://doi.org/10.1007/s00127-011-0406-4

Flouri, E., Midouhas, E., & Francesconi, M. (2020). Neighbourhood deprivation and child behaviour across childhood and adolescence. *Longitudinal and Life Course Studies*, *11*(2). https://doi.org/10.1332/175795919X15722477076216

Forero, D. A., Lopez-Leon, S., González-Giraldo, Y., & Bagos, P. G. (2019). Ten simple rules for carrying out and writing meta-analyses. In *PLoS Computational Biology* (Vol. 15, Issue 5). https://doi.org/10.1371/journal.pcbi.1006922

Forsythe, R., Murphy, C.-A., Tulip, J., & Law, J. (2020). Why Clinicians Choose Their Language Intervention Approach: An International Perspective on Intervention for Children with Developmental Language Disorder. *Folia Phoniatrica et Logopaedica*. https://doi.org/10.1159/000513242

Fraser, A., Macdonald-wallis, C., Tilling, K., Boyd, A., Golding, J., Davey smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A., Ring, S., Nelson, S. M., & Lawlor, D. A. (2013). Cohort profile: The avon longitudinal study of parents and children: ALSPAC mothers cohort. *International Journal of Epidemiology*, *42*(1). https://doi.org/10.1093/ije/dys066

Freeman, P. R., Hedges, L. V., & Olkin, I. (1986). Statistical Methods for Meta-Analysis. *Biometrics*, *42*(2). https://doi.org/10.2307/2531069

Freese, J. (2002). Least Likely Observations in Regression Models for Categorical

Outcomes. *The Stata Journal: Promoting Communications on Statistics and Stata*, *2*(3). https://doi.org/10.1177/1536867x0200200306

Frey, B. B. (2018). Post Hoc Analysis. In *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. https://doi.org/10.4135/9781506326139

Friesen, D. C., & Bialystok, E. (2012). Metalinguistic Ability in Bilingual Children: The Role of Executive Control. *Rivista Di Psicolinguistica Applicata*, *12*(3).

Frizelle, P., Tolonen, A. K., Tulip, J., Murphy, C. A., Saldana, D., & McKean, C. (2021a). The impact of intervention dose form on oral language outcomes for children with developmental language disorder. In *Journal of Speech, Language, and Hearing Research* (Vol. 64, Issue 8). https://doi.org/10.1044/2021_JSLHR-20-00734

Frizelle, P., Tolonen, A. K., Tulip, J., Murphy, C. A., Saldana, D., & McKean, C. (2021b). The influence of quantitative intervention dosage on oral language outcomes for children with developmental language disorder: A systematic review and narrative synthesis. In *Language, Speech, and Hearing Services in Schools* (Vol. 52, Issue 2). https://doi.org/10.1044/2020_LSHSS-20-00058

Fuchs, L. S., Gilbert, J. K., Fuchs, D., Seethaler, P. M., & Brittany, B. L. (2018). Text Comprehension and Oral Language as Predictors of Word-Problem Solving: Insights into Word-Problem Solving as a Form of Text Comprehension. *Scientific Studies of Reading*, *22*(2). https://doi.org/10.1080/10888438.2017.1398259

Gardner-Neblett, N., & Iruka, I. U. (2015). Oral narrative skills: Explaining the language-emergent literacy link by race/ethnicity and SES. *Developmental Psychology*, *51*(7). https://doi.org/10.1037/a0039274

Gathercole, V. M. (2018). Language Development in Bilingual Children: Fact, Factoid and Fiction. In *Language Policy(Netherlands)* (Vol. 15). https://doi.org/10.1007/978-3-319-75963-0_14

Geldsetzer, P., & Fawzi, W. (2017). Quasi-experimental study designs series—paper 2: complementary approaches to advancing global health knowledge. *Journal of Clinical Epidemiology*, *89*. https://doi.org/10.1016/j.jclinepi.2017.03.015

George, A., Hansen, K., & Schoon, I. (2007). Child Development. In K. Hansen & H.

Joshi (Eds.), *Millennium Cohort Study Second Survey: A User's Guide to Initial Findings* (pp. 94–109). Centre for Longitudinal Studies. https://cls.ucl.ac.uk/wp-content/uploads/2017/07/Users-Guide-to-Initial-Findings-20707.pdf

Gillam, R. B., Loeb, D. F., Hoffman, L. V. M., Bohman, T., Champlin, C. A., Thibodeau, L., Widen, J., Brandel, J., & Friel-Patti, S. (2008). The efficacy of Fast ForWord Language intervention in school-age children with language impairment: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research*, *51*(1). https://doi.org/10.1044/1092-4388(2008/007)

Glover, L., & Woods, M. (2020). *International Healthcare system profiles: Australia*. The Commonwealth Fund. https://www.commonwealthfund.org/international-health-policy-center/countries/australia

Goisis, A., Schneider, D. C., & Myrskylä, M. (2017). The reversing association between advanced maternal age and child cognitive ability: Evidence from three UK birth cohorts. *International Journal of Epidemiology*, *46*(3). https://doi.org/10.1093/ije/dyw354

Gombert, J. E. (1997). Metalinguistic Development in First-Language Acquisition. In *Encyclopedia of Language and Education*. https://doi.org/10.1007/978-94-011-4533-6_5

Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the strengths and difficulties questionnaire (SDQ): Data from british parents, teachers and children. *Journal of Abnormal Child Psychology*, *38*(8). https://doi.org/10.1007/s10802-010-9434-x

Gopalan, M., Rosinger, K., & Ahn, J. Bin. (2020). Use of Quasi-Experimental Research Designs in Education Research: Growth, Promise, and Challenges. *Review of Research in Education*, *44*(1). https://doi.org/10.3102/0091732X20903302

Gould, W., & Emeritus, P. (2011a). Merging data, part 1: Merges gone bad. *The Stata Blog*. https://blog.stata.com/2011/04/18/merging-data-part-1-merges-gone-bad/

Gould, W., & Emeritus, P. (2011b). Merging data, part 2: Multiple-key merges. In *the Stata blog*. https://blog.stata.com/2011/05/27/merging-data-part-2-multiple-key-merges/

Grant, S., Pedersen, E. R., Osilla, K. C., Kulesza, M., & D'Amico, E. J. (2016). It is time to develop appropriate tools for assessing minimal clinically important differences, performance bias and quality of evidence in reviews of behavioral interventions. In *Addiction* (Vol. 111, Issue 9). https://doi.org/10.1111/add.13380

Green, C. (2021). The oral language productive vocabulary profile of children starting school: A resource for teachers. *Australian Journal of Education*, *65*(1). https://doi.org/10.1177/0004944120982771

Greenwood, C. R., Schnitz, A. G., Carta, 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*. https://doi.org/10.1016/j.ecresq.2019.04.001

Gregg, P., Waldfogel, J., & Washbrook, E. (2006). Family expenditures post-welfare reform in the UK: Are low-income families starting to catch up? *Labour Economics*, *13*(6). https://doi.org/10.1016/j.labeco.2005.10.002

Gregoriadis, A., Tsigilis, N., Grammatikopoulos, V., & Kouli, O. (2016). Comparing quality of childcare and kindergarten centres: The need for a strong and equal partnership in the Greek early childhood education system. *Early Child Development and Care*, *186*(7). https://doi.org/10.1080/03004430.2015.1077820

Griffiths, S., Kievit, R. A., & Norbury, C. (2022). Mutualistic coupling of vocabulary and non-verbal reasoning in children with and without language disorder. *Developmental Science*. https://doi.org/10.1111/desc.13208

Grondhuis, S. N., Lecavalier, L., Arnold, L. E., Handen, B. L., Scahill, L., McDougle, C. J., & Aman, M. G. (2018). Differences in verbal and nonverbal IQ test scores in children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *49*. https://doi.org/10.1016/j.rasd.2018.02.001

Growing up in Australia. (2022). *The Longitudinal Study of Australian Children*. https://growingupinaustralia.gov.au/

Hahs-Vaughn, D. L., & Lomax, R. G. (2020). An Introduction to Statistical Concepts. In *An Introduction to Statistical Concepts*. https://doi.org/10.4324/9781315624358

Haidich, A. B. (2014). Meta-analysis in medical research Meta-analysis in medical research. *Hippokratia*, *14*(1).

Hammer, C. S., Morgan, P., Farkas, G., Hillemeier, M., Bitetti, D., & Maczuga, S. (2017). Late talkers: A population-based study of risk factors and school readiness consequences. *Journal of Speech, Language, and Hearing Research*, *60*(3). https://doi.org/10.1044/2016_JSLHR-L-15-0417

Handley, M. A., Lyles, C. R., McCulloch, C., & Cattamanchi, A. (2018). Selecting and Improving Quasi-Experimental Designs in Effectiveness and Implementation Research. *Annual Review of Public Health*, *39*, 5–25. https://doi.org/10.1146/annurev-publhealth-040617-014128

Hansen, K., & Jones, E. M. (2008). Foundation stage profile and devolved administration teacher survey. In *Millennium Cohort Study Third Survey: A User's Guide to Initial Findings* (pp. 98–117). https://cls.ucl.ac.uk/wp-content/uploads/2017/07/MCS-3-Descriptive-Report-Oct-2008.pdf

Hansen, K, Joshi, H., & Dex, S. (2010). *Children of the 21st Century: The first five years* (K Hansen, H. Joshi, & S. Dex (eds.)). The Policy Press.

Hansen, Kirstine, & Kneale, D. (2013). Does How You Measure Income Make a Difference to Measuring Poverty? Evidence from the UK. *Social Indicators Research*, *110*(3). https://doi.org/10.1007/s11205-011-9976-5

Harding, J. F., Morris, P. A., & Hughes, D. (2015). The Relationship Between Maternal Education and Children's Academic Outcomes: A Theoretical Framework. *Journal of Marriage and Family*, *77*(1). https://doi.org/10.1111/jomf.12156

Haskill, A. M., & Tyler, A. A. (2007). A comparison of linguistic profiles in subgroups of children with specific language impairment. *American Journal of Speech-Language Pathology*, *16*(3). https://doi.org/10.1044/1058-0360(2007/026)

Hayes, A. F. (2017). Introduction to Mediation, Moderation, and Conditional Process Analysis, Second Edition: A Regression-Based Approach. In *the Guilford Press* (Issue 3).

Hedges L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press. https://doi.org/10.1016/C2009-0-03396-0

Heesen, M., Klimek, M., & Hoeks, S. E. (2018). Restrictive or responsive? Outcome classification and unplanned sub-group analyses in meta-analyses. In *Anaesthesia* (Vol. 73, Issue 3). https://doi.org/10.1111/anae.14078

Herbert, R. D. (2000). How to estimate treatment effects from reports of clinical trials. I: Continuous outcomes. *Australian Journal of Physiotherapy*, *46*(3). https://doi.org/10.1016/S0004-9514(14)60334-2

Higgins, J., Deeks, J., & Altman, D. (2021). General principles for dealing with missing data. In J. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, & V. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane.

Higgins, J. P. T., Altman, D. G., & Sterne, J. A. C. (2018). Assessing risk of bias in included studies. In J. P. T. Higgins & S. Green (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane Statistical Methods Group and the Cochrane Bias Methods Group. https://handbook-5-1.cochrane.org/chapter_8/8_assessing_risk_of_bias_in_included_studies.htm

Hiilamo, H., Merikukka, M., & Haataja, A. (2018). Long-Term Educational Outcomes of Child Care Arrangements in Finland. *SAGE Open*, *8*(2). https://doi.org/10.1177/2158244018774823

Hobcraft, J. N., & Kiernan, K. E. (2010). *Predictive factors from age 3 and infancy for poor child outcomes at age 5 relating to children's development, behaviour and health: evidence from the Millennium Cohort Study.* .

Hoff, E. (2013). *Language Development*. Wadsworth, Cengage Learning.

Honig, A. S. (2007). Oral language development. *Early Child Development and Care*, *117*(6–7), 581–613.

Hopkin, R., Stokes, L., & Wilkinson, D. (2009). Using foundation stage profile assessments to assess outcomes from early years education. *National Institute Economic Review*, *207*(1). https://doi.org/10.1177/0027950109103691

Hosokawa, R., & Katsura, T. (2018). Effect of socioeconomic status on behavioral problems from preschool to early elementary school – A Japanese longitudinal study. *PLoS ONE*, *13*(5). https://doi.org/10.1371/journal.pone.0197961

ICAN. (2021). *Welcome to "What works?"* ICAN. https://ican.org.uk/i-cans-talking-point/professionals/tct-resources/what-works-database/

Illøkken, K. E., Johannessen, B., Barker, M. E., Hardy-Johnson, P., Øverby, N. C., & Vik, F. N. (2021). Free school meals as an opportunity to target social equality, healthy eating, and school functioning: Experiences from students and teachers in Norway. *Food and Nutrition Research*, *65*. https://doi.org/10.29219/FNR.V65.7702

Isaacs, J. B. (2012). *Starting School at a Disadvantage: The School Readiness of Poor Children. The Social Genome Project.* https://www.brookings.edu/wpcontent/uploads/2016/06/0319_school_disadvantage_isaacs.pdf.

Jackson, M. I., Kiernan, K., & McLanahan, S. (2017). Maternal Education, Changing Family Circumstances, and Children's Skill Development in the United States and UK. *Annals of the American Academy of Political and Social Science*, *674*(1). https://doi.org/10.1177/0002716217729471

Jiang, H., Logan, J. A., Jia, R., Justice, L. M., Lomax, R., O'connell, A., Pentimonti, J., Petrill, S. A., Piasta, S. B., Gray, S., Restrepo, M. A., Cain, K., Catts, H., Bridges, M., Nielsen, D., Hogan, T., Bovaird, J., & Nelson, J. R. (2018). Modeling the nature of grammar and vocabulary trajectories from prekindergarten to third grade. *Journal of Speech, Language, and Hearing Research*, *61*(4). https://doi.org/10.1044/2018_JSLHR-L-17-0090

Jimenez, M. E., Wade, R., Lin, Y., Morrow, L. M., & Reichman, N. E. (2016). Adverse experiences in early childhood and kindergarten outcomes. *Pediatrics*, *137*(2). https://doi.org/10.1542/peds.2015-1839

Johnson, C. J., Beitchman, J. H., & 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*, *9*, 51–65. https://doi.org/10.1044/1058-0360(2009/08-0083)

Johnson, J. (2008). *Millennium Third Survey Follow-up: A Guide to the School Assessment Datasets*. https://cls.ucl.ac.uk/wp-content/uploads/2017/07/Guide-to-the-School-Assessment-Datasets-1.pdf

Johnson, T., Burgoyne, A. P., Mix, K. S., Young, C. J., & Levine, S. C. (2022). Spatial and mathematics skills: Similarities and differences related to age, SES, and gender. *Cognition*, *218*. https://doi.org/10.1016/j.cognition.2021.104918

Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health*, *105*(11). https://doi.org/10.2105/AJPH.2015.302630

Joshi, H., & Fitzsimons, E. (2016). The Millennium Cohort Study: The making of a multi-purpose resource for social science and policy. *Longitudinal and Life Course Studies*, *4*(4), 409–430.

Joshi, P., Geronimo, K., & Acevedo-Garcia, D. (2016). Head start since the war on poverty: Taking on new challenges to address persistent school readiness gaps. In *Journal of Applied Research on Children* (Vol. 7, Issue 1).

Junaid, K. A., & Fellowes, S. (2006). Gender differences in the attainment of motor skills on the movement assessment battery for children. *Physical and Occupational Therapy in Pediatrics*, *26*(1–2). https://doi.org/10.1300/J006v26n01_02

Kabisch, M., Ruckes, C., Seibert-Grafe, M., & Blettner, M. (2011). Randomized controlled trials: Part 17 of a series on evaluation of scientific publications. *Deutsches Arzteblatt International*, *108*(39). https://doi.org/10.3238/arztebl.2011.0663

Kahlert, J., Gribsholt, S. B., Gammelager, H., Dekkers, O. M., & Luta, G. (2017). Control of confounding in the analysis phase – an overview for clinicians. *Clinical Epidemiology*, *9*. https://doi.org/10.2147/CLEP.S129886

Kasza, J., & Wolfe, R. (2014). Interpretation of commonly used statistical regression models. In *Respirology* (Vol. 19, Issue 1). https://doi.org/10.1111/resp.12221

Kay, L. (2018). School Readiness: A Culture of Compliance? *The University of Sheffield*, *February*.

Kiernan, K. E., & Mensah, F. K. (2009). Poverty, maternal depression, family status and children’s cognitive and behavioural development in early childhood: A longitudinal study. *Journal of Social Policy*, *38*(4). https://doi.org/10.1017/S0047279409003250

Kim, H.-Y. (2018). Statistical notes for clinical researchers: covariance and correlation. *Restorative Dentistry & Endodontics*, *43*(1). https://doi.org/10.5395/rde.2018.43.e4

Kim, Y.-S. G., & Pilcher, H. (2016). *What Is Listening Comprehension and What Does It Take to Improve Listening Comprehension?* https://doi.org/10.1007/978-3-319-31235-4_10

King, T., McKean, C., Rush, R., Westrupp, E. M., Mensah, F. K., Reilly, S., & Law, J. (2017). Acquisition of maternal education and its relation to single-word reading in middle childhood: An analysis of the millennium cohort study. *Merrill-Palmer Quarterly*, *63*(2). https://doi.org/10.13110/merrpalmquar1982.63.2.0181

Klem, M., Hagtvet, B., Hulme, C., & Gustafsson, J. E. (2016). Screening for language delay: Growth trajectories of language ability in low- and high-performing children. *Journal of Speech, Language, and Hearing Research*, *59*(5), 1035–1045. https://doi.org/10.1044/2016_JSLHR-L-15-0289

Kluczniok, K., Lehrl, S., Kuger, S., & Rossbach, H. G. (2013). Quality of the home learning environment during preschool age - Domains and contextual conditions. *European Early Childhood Education Research Journal*, *21*(3). https://doi.org/10.1080/1350293X.2013.814356

Knapp, T. R. (2016). Why Is the One-Group Pretest–Posttest Design Still Used? In *Clinical Nursing Research* (Vol. 25, Issue 5). https://doi.org/10.1177/1054773816666280

Kokkalia, G., Drigas, A., Economou, A., & Roussos, P. (2019). School readiness from kindergarten to primary school. *International Journal of Emerging Technologies in Learning*, *14*(11). https://doi.org/10.3991/IJET.V14I11.10090

Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, *49*(4). https://doi.org/10.3102/0013189X20912798

Lange, B. P., Euler, H. A., & Zaretsky, E. (2016). Sex differences in language competence of 3- to 6-year-old children. *Applied Psycholinguistics*, *37*(6). https://doi.org/10.1017/S0142716415000624

Law, J. (2015). *The importance of oral language and its implications for early years practice: A report to Goodstart Early Learning*.

https://www.goodstart.org.au/getmedia/8cd15601-203c-441a-a54f-61caccb4eb22/The-importance-of-oral-language-and-its-implications-for-early-years-practice.pdf.aspx

Law, J., Tulip, J., & Beckermann, E. (2019). The development of the practitioner survey. In *Managing Children with Developmental Language Disorder: Theory and Practice across Europe and Beyond*. https://doi.org/10.4324/9780429455308

Law, J, Charlton, J., Dockrell, J., Gascoigne, M., McKean, C., & Theakston, A. (2017). *Early Language Development: Needs, provision and intervention for pre-school children from socio-economically disadvantaged backgrounds.* https://eprints.ncl.ac.uk/file_store/production/242270/99F15C18-813F-4447-BE90-0BBDAA211E0C.pdf

Law, J., Charlton, J., Mckean, C., Beyer, F., Fernandez-Garcia, C., Mashayekhi, A., & Rush, R. (2018). Parent-child reading to improve language development and school readiness: A systematic review and meta-analysis. *Newcastle University EPrints*, *July*. https://eprints.ncl.ac.uk/file_store/production/246226/B804D16A-F0A4-43C0-920F-4A274A131AAF.pdf

Law, J., Dennis, J. A., & Charlton, J. J. V. (2017). Speech and language therapy interventions for children with primary speech and/or language disorders. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD012490

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*. https://doi.org/10.1044/1092-4388(2004/069)

Law, J., Lee, W., Lindsay, G., Roulstone, S., Wren, Y., & Zeng, B. (2012). *“What Works”: interventions for children and young people with speech, language and communication needs*. https://dera.ioe.ac.uk//16323/

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

Law, J., & Stringer, H. (2014). The overlap between behaviour and communication and its implications for mental health in childhood: the elephant in the room. *Emotional and Behavioural Difficulties*, *19*(1). https://doi.org/10.1080/13632752.2013.854959

Lee, K., & Zhang, L. (2021). Cumulative Effects of Poverty on Children's Social-Emotional Development: Absolute Poverty and Relative Poverty. *Community Mental Health Journal*. https://doi.org/10.1007/s10597-021-00901-x

Lee, Y. H. (2018). An overview of meta-analysis for clinicians. In *Korean Journal of Internal Medicine* (Vol. 33, Issue 2). https://doi.org/10.3904/kjim.2016.195

Leech, K., Wei, R., Harring, J. R., & Rowe, M. L. (2018). A brief parent-focused intervention to improve preschoolers' conversational skills and school readiness. *Developmental Psychology*, *54*(1). https://doi.org/10.1037/dev0000411

Letourneau, N. L., Duffett-Leger, L., Levac, L., Watson, B., & Young-Morris, C. (2013). Socioeconomic Status and Child Development: A Meta-Analysis. *Journal of Emotional and Behavioral Disorders*, *21*(3). https://doi.org/10.1177/1063426611421007

Levickis, P., Sciberras, E., McKean, C., Conway, L., Pezic, A., Mensah, F. K., Bavin, E. L., Bretherton, L., Eadie, P., Prior, M., & Reilly, S. (2018). Language and social-emotional and behavioural wellbeing from 4 to 7 years: a community-based study. *European Child and Adolescent Psychiatry*, *27*(7). https://doi.org/10.1007/s00787-017-1079-7

Lewis, B. A., Freebairn, L., Tag, J., Ciesla, A. A., Iyengar, S. K., Stein, C. M., & Taylor, H. G. (2015). Adolescent outcomes of children with early speech sound disorders with and without language impairment. *American Journal of Speech-Language Pathology*, *24*(2), 150–163. https://doi.org/10.1044/2014_AJSLP-14-0075

Lewis, K., Sandilos, L. E., Hammer, C. S., Sawyer, B. E., & Méndez, L. I. (2016). Relations Among the Home Language and Literacy Environment and Children's Language Abilities: A Study of Head Start Dual Language Learners and Their Mothers. *Early Education and Development*, *27*(4). https://doi.org/10.1080/10409289.2016.1082820

Li, W., Farkas, G., Duncan, G. J., Burchinal, M. R., & Vandell, D. L. (2013). Timing of high-quality child care and cognitive, language, and preacademic development. *Developmental Psychology*, *49*(8). https://doi.org/10.1037/a0030613

Linden, A. (2013). Assessing regression to the mean effects in health care initiatives. *BMC Medical Research Methodology*, *13*(1). https://doi.org/10.1186/1471-2288-13-119

Lindsay, G., Wedell, K., & Dockrell, J. (2020). Warnock 40 Years on: The Development of Special Educational Needs Since the Warnock Report and Implications for the Future. *Frontiers in Education*, *4*. https://doi.org/10.3389/feduc.2019.00164

Lobo, M. A., Moeyaert, M., Cunha, A. B., & Babik, I. (2017). Single-case design, analysis, and quality assessment for intervention research. *Journal of Neurologic Physical Therapy*, *41*(3). https://doi.org/10.1097/NPT.0000000000000187

Locke, A., Ginsborg, J., & Peers, I. (2002). Development and disadvantage: Implications for the early years and beyond. . *International Journal of Language and Communication Disorders* , *37*, 657–664.

Lonigan, C. J., & Milburn, T. F. (2017). Identifying the dimensionality of oral language skills of children with typical development in preschool through fifth grade. *Journal of Speech, Language, and Hearing Research*, *60*(8). https://doi.org/10.1044/2017_JSLHR-L-15-0402

Lonigan, C. J., Phillips, B. M., Clancy, J. L., Landry, S. H., Swank, P. R., Assel, M., Taylor, H. B., Klein, A., Starkey, P., Domitrovich, C. E., Eisenberg, N., de Villiers, J., de Villiers, P., & Barnes, M. (2015). Impacts of a Comprehensive School Readiness Curriculum for Preschool Children at Risk for Educational Difficulties. *Child Development*, *86*(6). https://doi.org/10.1111/cdev.12460

Lovas, G. S. (2011). Gender and patterns of language development in mother-toddler and father-toddler dyads. *First Language*, *31*(1). https://doi.org/10.1177/0142723709359241

Magnuson, K. A., Sexton, H. R., Davis-Kean, P. E., & Huston, A. C. (2009). Increases in maternal education and young children’s language skills. *Merrill-Palmer Quarterly*, *55*(3). https://doi.org/10.1353/mpq.0.0024

Maguire, L. K., Niens, U., McCann, M., & Connolly, P. (2016). Emotional development among early school-age children: gender differences in the role of problem behaviours. *Educational Psychology*, *36*(8). https://doi.org/10.1080/01443410.2015.1034090

Marinis, T., Armon-Lotem, S., & Pontikas, G. (2017). Language impairment in bilingual children: State of the art 2017. *Linguistic Approaches to Bilingualism*, *7*(3–4). https://doi.org/10.1075/lab.00001.mar

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young Children’s word learning: A meta-analysis. *Review of Educational Research*, *80*(3), 300–335. https://doi.org/10.3102/0034654310377087

Marulis, L. M., & Neuman, S. B. (2013). How Vocabulary Interventions Affect Young Children at Risk: A Meta-Analytic Review. *Journal of Research on Educational Effectiveness*, *6*(3), 223–262. https://doi.org/10.1080/19345747.2012.755591

Masnjak, M. (2017). Gender differences in social emotional development and physical activity level in preschool children. *Proceedings of the 8th International Scientific Conference on Kinesiology*, 530–534.

Matz, S. C., Gladstone, J. J., & Stillwell, D. (2017). In a World of Big Data, Small Effects Can Still Matter: A Reply to Boyce, Daly, Hounkpatin, and Wood (2017). In *Psychological Science* (Vol. 28, Issue 4). https://doi.org/10.1177/0956797617697445

Maurice-Stam, H., Haverman, L., Splinter, A., van Oers, H. A., Schepers, S. A., & Grootenhuis, M. A. (2018). Dutch norms for the Strengths and Difficulties Questionnaire (SDQ) - parent form for children aged 2-18years. *Health and Quality of Life Outcomes*, *16*(1). https://doi.org/10.1186/s12955-018-0948-1

McDonald, S. W., Kehler, H. L., & Tough, S. C. (2016). Protective factors for child development at age 2 in the presence of poor maternal mental health: Results from the All Our Babies (AOB) pregnancy cohort. *BMJ Open*, *6*(11). https://doi.org/10.1136/bmjopen-2016-012096

McDowell, K. D., Lonigan, C. J., & Goldstein, H. (2007). Relations among socioeconomic status, age, and predictors of phonological awareness. *Journal of Speech, Language, and Hearing Research*, *50*(4). https://doi.org/10.1044/1092-4388(2007/075)

McGauran, N., Wieseler, B., Kreis, J., Schüler, Y. B., Kölsch, H., & Kaiser, T. (2010). Reporting bias in medical research - a narrative review. In *Trials* (Vol. 11). https://doi.org/10.1186/1745-6215-11-37

McGrath, L. M., Hutaff-Lee, C., Scott, A., Boada, R., Shriberg, L. D., & Pennington, B. F. (2008). Children with comorbid speech sound disorder and specific language impairment are at increased risk for attention-deficit/hyperactivity disorder. *Journal of Abnormal Child Psychology*, *36*(2). https://doi.org/10.1007/s10802-007-9166-8

McKean, C., Gerrits, E., Tulip, J., & Tolonen, A.-K. (2019). Service delivery for children with language disorders across europe and beyond. In *Managing Children with Developmental Language Disorder: Theory and Practice across Europe and Beyond*. https://doi.org/10.4324/9780429455308

McKean, C., Mensah, F. K., Eadie, P., Bavin, E. L., Bretherton, L., Cini, E., & Reilly, S. (2015). Levers for language growth: Characteristics and predictors of language trajectories between 4 and 7 years. *PLoS ONE*, *10*(8). https://doi.org/10.1371/journal.pone.0134251

McKean, C., Wraith, D., Eadie, P., Cook, F., Mensah, F., & Reilly, S. (2017). 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 and Allied Disciplines*, *58*(10). https://doi.org/10.1111/jcpp.12790

McKenzie, J. E., Brennan, S. E., Ryan, R. E., Thomson, H. J., Johnston, R. V., & Thomas, J. (2021). Defining the criteria for including studies and how they will be grouped for the synthesis. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions*. https://training.cochrane.org/handbook/current/chapter-03

McLeod, S., & McKinnon, D. H. (2007). Prevalence of communication disorders compared with other learning needs in 14 500 primary and secondary school students. *International Journal of Language and Communication Disorders*, *42*(SUPPL. 1). https://doi.org/10.1080/13682820601173262

Melhuish, E. C., Phan, M. B., Sylva, K., Sammons, P., Siraj-Blatchford, I., & Taggart,

B. (2008). Effects of the home learning environment and preschool center experience upon literacy and numeracy development in early primary school. *Journal of Social Issues*, *64*(1). https://doi.org/10.1111/j.1540-4560.2008.00550.x

Melhuish, E., & Gardiner, J. (2018). *Study of Early Education and Development (SEED): Study of Quality of Early Years Provision in England (Revised): Research Report*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/723736/Study_of_quality_of_early_years_provision_in_England.pdf

Mensah, F. K., & Kiernan, K. E. (2010a). Gender differences in educational attainment: Influences of the family environment. *British Educational Research Journal*, *36*(2). https://doi.org/10.1080/01411920902802198

Mensah, F. K., & Kiernan, K. E. (2010b). Parents' mental health and children's cognitive and social development: Families in England in the Millennium Cohort Study. *Social Psychiatry and Psychiatric Epidemiology*, *45*(11). https://doi.org/10.1007/s00127-009-0137-y

Miller, C. J., Smith, S. N., & Pugatch, M. (2020). Experimental and quasi-experimental designs in implementation research. *Psychiatry Research*, *283*. https://doi.org/10.1016/j.psychres.2019.06.027

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., Estarli, M., Barrera, E. S. A., Martínez-Rodríguez, R., Baladia, E., Agüero, S. D., Camacho, S., Buhring, K., Herrero-López, A., Gil-González, D. M., Altman, D. G., Booth, A., … Whitlock, E. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), 1–9. https://doi.org/10.1186/2046-4053-4-1

Mol, S. E., & 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), 267–296. https://doi.org/10.1037/a0021890

Mollborn, S., Lawrence, E., James-Hawkins, L., & Fomby, P. (2014). When do socioeconomic resources matter most in early childhood? *Advances in Life Course Research*, *20*. https://doi.org/10.1016/j.alcr.2014.03.001

Montoya, A. K. (2019). Moderation analysis in two-instance repeated measures designs: Probing methods and multiple moderator models. *Behavior Research Methods*, *51*(1). https://doi.org/10.3758/s13428-018-1088-6

Montroy, J. J., Merz, E. C., Williams, J. M., Landry, S. H., Johnson, U. Y., Zucker, T. A., Assel, M., Taylor, H. B., Lonigan, C. J., Phillips, B. M., Clancy-Menchetti, J., Barnes, M. A., Eisenberg, N., Spinrad, T., Valiente, C., de Villiers, J., & de Villiers, P. (2019). Hot and cool dimensionality of executive function: Model invariance across age and maternal education in preschool children. *Early Childhood Research Quarterly*, *49*. https://doi.org/10.1016/j.ecresq.2019.06.011

Morinis, J., Carson, C., & Quigley, M. A. (2013). Effect of teenage motherhood on cognitive outcomes in children: A population-based cohort study. *Archives of Disease in Childhood*, *98*(12). https://doi.org/10.1136/archdischild-2012-302525

Murray, E., & Harrison, L. J. (2011). The influence of being ready to learn on children's early school literacy and numeracy achievement. *Educational Psychology*, *31*(5). https://doi.org/10.1080/01443410.2011.573771

Murray, E., Thomas, D., & McKechnie, J. (2019). Comorbid morphological disorder apparent in some children aged 4-5 years with childhood apraxia of speech: findings from standardised testing. *Clinical Linguistics and Phonetics*, *33*(1–2). https://doi.org/10.1080/02699206.2018.1513565

National Institute for Health and Care Excellence. (2021). *Antisocial behaviour and conduct disorders in children and young people: recognition and management*. https://www.nice.org.uk/guidance/cg158/chapter/introduction

National Institute of Child Health and Human Development. (2006). *The NICHD study of early child care and youth development*. https://www.nichd.nih.gov/sites/default/files/publications/pubs/documents/seccyd_06.pdf

NCES. (2022). *The Early Childhood Longitudinal Study*. https://nces.ed.gov/ecls/

Neuman, S. B., Kaefer, T., & Pinkham, A. M. (2018). Double dose of disadvantage: Language experiences for low-income children in home and school. *Journal of Educational Psychology*, *110*(1). https://doi.org/10.1037/edu0000201

Nicholls, M., Neale, I., Joyner, O., & Sheikh, M. (2020). *Kindred2 –School Readiness*.

Nijhof, S. L., Vinkers, C. H., van Geelen, S. M., Duijff, S. N., Achterberg, E. J. M., van der Net, J., Veltkamp, R. C., Grootenhuis, M. A., van de Putte, E. M., Hillegers, M. H. J., van der Brug, A. W., Wierenga, C. J., Benders, M. J. N. L., Engels, R. C. M. E., van der Ent, C. K., Vanderschuren, L. J. M. J., & Lesscher, H. M. B. (2018). Healthy play, better coping: The importance of play for the development of children in health and disease. In *Neuroscience and Biobehavioral Reviews* (Vol. 95). https://doi.org/10.1016/j.neubiorev.2018.09.024

Niklas, F., & Schneider, W. (2017). Home learning environment and development of child competencies from kindergarten until the end of elementary school. *Contemporary Educational Psychology*, *49*. https://doi.org/10.1016/j.cedpsych.2017.03.006

Nix, R. L., Bierman, K. L., Domitrovich, C. E., & Gill, S. (2013). Promoting preschool social emotional skills with the Head Start REDI Program enhances academic and behavioral outcomes in kindergarten. *Early Education and Development*, *24*, 1000–1019.

Noble, K. G., Duch, H., Darvique, M. E., Grundleger, A., Rodriguez, C., & Landers, C. (2012). “Getting Ready for School:” A Preliminary Evaluation of a Parent-Focused School-Readiness Program. *Child Development Research*, *2012*. https://doi.org/10.1155/2012/259598

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 and Psychiatry*, *11*, 1247–1257. https://doi.org/10.1111/jcpp.12573

Norbury, C. F., Vamvakas, G., Gooch, D., Baird, G., Charman, T., Simonoff, E., & Pickles, A. (2017). Language growth in children with heterogeneous language disorders: a population study. *Journal of Child Psychology and Psychiatry*, *58*(10), 1092–1105. https://doi.org/10.1111/jcpp.12793

Office for Standards in Education, C. S. and S. (2014). *Are you ready?: Good practice in school readiness*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/att achment_data/file/418819/Are_you_ready_Good_practice_in_school_readiness.pdf

Organisation for Economic Co-operation and Development. (2010). The high cost of low educational performance: The long-run economic impact of improving PISA outcomes. In *OECD Publishing*. https://www.oecd.org/pisa/44417824.pdf

Organisation for Economic Co-operation and Development. (2012). *OECD project on the distribution of household incomes*.

Organisation for Economic Co-operation and Development. (2021). *Poverty rate*. https://doi.org/10.1787/0fe1315d-en

Orri, M., Boivin, M., Chen, C., Ahun, M. N., Geoffroy, M. C., Ouellet-Morin, I., Tremblay, R. E., & Côté, S. M. (2021). Cohort Profile: Quebec Longitudinal Study of Child Development (QLSCD). *Social Psychiatry and Psychiatric Epidemiology*, *56*(5). https://doi.org/10.1007/s00127-020-01972-z

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

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. In *The BMJ* (Vol. 372). https://doi.org/10.1136/bmj.n71

Page, M. J., Shamseer, L., & Tricco, A. C. (2018). Registration of systematic reviews in PROSPERO: 30,000 records and counting. *Systematic Reviews*, *7*(1). https://doi.org/10.1186/s13643-018-0699-4

Pallier, C. (2007). *Critical periods in language acquisition and language attrition*. https://doi.org/10.1075/sibil.33.11pal

Pan, Q., Trang, K. T., Love, H. R., & Templin, J. (2019). School Readiness Profiles and Growth in Academic Achievement. *Frontiers in Education*, *4*. https://doi.org/10.3389/feduc.2019.00127

Papachristou, E., & Flouri, E. (2020). The codevelopment of internalizing symptoms, externalizing symptoms, and cognitive ability across childhood and adolescence. *Development and Psychopathology*, *32*(4). https://doi.org/10.1017/S0954579419001330

Patalay, P., Fink, E., Fonagy, P., & Deighton, J. (2016). Unpacking the associations between heterogeneous externalising symptom development and academic attainment in middle childhood. *European Child and Adolescent Psychiatry*, *25*(5). https://doi.org/10.1007/s00787-015-0758-5

Price, P.C., Jhangiani, R.S., & I.-Chant, A. C. (2019). Reliability and Validity of Measurement. In Jhangiani, R.S., I-Chant, A., Chiang, C.C., and Leighton, D.C. (Eds.) *Research Methods in Psychology*. https://kpu.pressbooks.pub/psychmethods4e/chapter/reliability-and-validity-of-measurement/

Paul, R. (2020). Language disorders. *Handbook of clinical neurology, 174*, 21-35. https://www.sciencedirect.com/science/article/abs/pii/B9780444641489000028?via%3Dihub.

Paull, G. ., Taylor, J., Duncan, A., & Payne, J. (2002). *Mothers' employment and childcare use in Britain*. Institute for Fiscal Studies. https://ifs.org.uk/comms/r64.pdf

Pellicano, E., Kenny, L., Brede, J., Klaric, E., Lichwa, H., & McMillin, R. (2017). Executive function predicts school readiness in autistic and typical preschool children. *Cognitive Development*, *43*. https://doi.org/10.1016/j.cogdev.2017.02.003

Peña, E. D., Bedore, L. M., Lugo-Neris, M. J., & Albudoor, N. (2020). Identifying Developmental Language Disorder in School Age Bilinguals: Semantics, Grammar, and Narratives. *Language Assessment Quarterly*, *17*(5). https://doi.org/10.1080/15434303.2020.1827258

Perrin, H. T., Feldman, H. M., & Huffman, L. C. (2020). Development and Evaluation of a School Readiness Curriculum for Pediatrics Residents. In *MedEdPORTAL : the journal of teaching and learning resources* (Vol. 16). https://doi.org/10.15766/mep_2374-8265.10976

Plewis, I. (2007). The Millennium Cohort Study: Technical Report on Sampling. In *Centre for Longitudinal Studies*.

http://doc.ukdataservice.ac.uk/doc/4683/mrdoc/pdf/mcs_technical_report_on_sampling_4th_edition.pdf

Pretti-Frontczak, K. (2014). Stop Trying to Make Kids “Ready” for Kindergarten. *Young Exceptional Children*, *17*(1). https://doi.org/10.1177/1096250614523346

Prior, M., Bavin, E., & Ong, B. (2011). Predictors of school readiness in five-to six-year-old children from an Australian longitudinal community sample. *Educational Psychology*, *31*(1). https://doi.org/10.1080/01443410.2010.541048

Igartua, J. J., & Hayes, A. F. (2021). Mediation, moderation, and conditional process analysis: Concepts, computations, and some common confusions. *The Spanish Journal of Psychology*, *24*.Quigley, M. A., Poulsen, G., Boyle, E., Wolke, D., Field, D., Alfirevic, Z., & Kurinczuk, J. J. (2012). Early term and late preterm birth are associated with poorer school performance at age 5 years: A cohort study. *Archives of Disease in Childhood: Fetal and Neonatal Edition*, *97*(3). https://doi.org/10.1136/archdischild-2011-300888

Racine, N., Plamondon, A., Madigan, S., McDonald, S., & Tough, S. (2018). Maternal adverse childhood experiences and infant development. *Pediatrics*, *141*(4). https://doi.org/10.1542/peds.2017-2495

Rahman, M. A., Todd, C., John, A., Tan, J., Kerr, M., Potter, R., Kennedy, J., Rice, F., & Brophy, S. (2018). School achievement as a predictor of depression and self-harm in adolescence: Linked education and health record study. *British Journal of Psychiatry*, *212*(4). https://doi.org/10.1192/bjp.2017.69

Redmond, S. M. (2016). Language impairment in the attention-deficit/hyperactivity disorder context. In *Journal of Speech, Language, and Hearing Research* (Vol. 59, Issue 1). https://doi.org/10.1044/2015_JSLHR-L-15-0038

Redmond, S. M., Ash, A. C., & Hogan, T. P. (2015). Consequences of co-occurring attention-deficit/Hyperactivity disorder on children’s language impairments. *Language, Speech, and Hearing Services in Schools*, *46*(2). https://doi.org/10.1044/2014_LSHSS-14-0045

Reid, L. D., & Strobino, D. M. (2019). A Population-Based Study of School Readiness Determinants in a Large Urban Public School District. *Maternal and Child Health Journal*, *23*(3). https://doi.org/10.1007/s10995-018-2666-z

Reilly, S., Cook, F., Bavin, E. L., Bretherton, L., Cahir, P., Eadie, P., Gold, L., Mensah, F., Papadopoullos, S., & Wake, M. (2018). Cohort profile: the early language in victoria study (ELVS). *International Journal of Epidemiology.*, *47*(1), 11–20.

Reilly, Sheena, Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., Goldfeld, S., Nicholson, J. M., & Wake, M. (2014). Specific language impairment: A convenient label for whom? In *International Journal of Language and Communication Disorders* (Vol. 49, Issue 4). https://doi.org/10.1111/1460-6984.12102

Reynolds, A. J., Temple, J. A., Ou, S. R., Arteaga, I. A., & White, B. A. B. (2011). School-based early childhood education and age-28 well-being: Effects by timing, dosage, and subgroups. *Science*, *333*(6040). https://doi.org/10.1126/science.1203618

Ricciardi, C., Manfra, L., Hartman, S., Bleiker, C., Dineheart, L., & Winsler, A. (2021). School readiness skills at age four predict academic achievement through 5th grade. *Early Childhood Research Quarterly*, *57*. https://doi.org/10.1016/j.ecresq.2021.05.006

Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific Language impairment: A longitudinal study from 2;6 to 21 years of age. *Journal of Speech, Language, and Hearing Research*, *58*(2). https://doi.org/10.1044/2015_JSLHR-L-14-0150

Roberts-Holmes, G. (2019). School readiness, governance and early years ability grouping. *Contemporary Issues in Early Childhood*, *22*(3). https://doi.org/10.1177/1463949119863128

Roberts, M. Y., & Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *American Journal of Speech-Language Pathology*, *20*, 180–199. https://doi.org/10.1044/1058-0360(2011/10-0055)

Rodriguez, E. T., & Tamis-LeMonda, C. S. (2011). Trajectories of the home learning environment across the first 5 years: Associations with children’s vocabulary and literacy skills at prekindergarten. *Child Development*, *82*(4), 1058–1075.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments.* John Wiley and Sons.

Roy, P., & Chiat, S. (2014). Developmental pathways of language and social communication problems in 9-11 year olds: Unpicking the heterogeneity. *Research in Developmental Disabilities*, *35*(10). https://doi.org/10.1016/j.ridd.2014.06.014

Royal College of Paediatrics and Child Health. (2015). *Facing the Future: Together for Child Health.* https://www.rcpch.ac.uk/sites/default/files/Facing_the_Future_Together_for_Child_Health.pdf

Royal College of Paediatrics and Child Health. (2021). *Long Term Conditions*. https://stateofchildhealth.rcpch.ac.uk/evidence/long-term-conditions/

Russo, J. M., Williford, A. P., Markowitz, A. J., Vitiello, V. E., & Bassok, D. (2019). Examining the validity of a widely-used school readiness assessment: Implications for teachers and early childhood programs. *Early Childhood Research Quarterly*, *48*. https://doi.org/10.1016/j.ecresq.2019.02.003

Ryan, R. (2013). Cochrane Consumers and Communication Review Group: data synthesis and analysis. *Cochrane Consumers and Communication Review Group*, *2013*(June).

Sadler, K., Akister, J., & Burch, S. (2015). Who are the young people who are not in education, employment or training? An application of the risk factors to a rural area in the UK. *International Social Work*, *58*(4). https://doi.org/10.1177/0020872813515010

Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., & Boyce, W. T. (2011). Family socioeconomic status and child executive functions: The roles of language, home environment, and single parenthood. *Journal of the International Neuropsychological Society*, *17*(1). https://doi.org/10.1017/S1355617710001335

Schleppegrell, M. J. (2012). Academic Language in Teaching and Learning: Introduction to the Special Issue. *The Elementary School Journal*, *112*(3).

Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. In *Journal of Clinical Epidemiology* (Vol. 98). https://doi.org/10.1016/j.jclinepi.2017.12.006

Schreiber-Gregory, D., & Bader, K. (2018). Logistic and Linear Regression Assumptions: Violation Recognition and Control. *Midwest SAS User Group*.

Schünemann, H. J., Vist, G. E., Higgins, J. P. T., Santesso, N., Deeks, J. J., Glasziou, P., Akl, E. A., & Guyatt, G. H. (2019). Interpreting results and drawing conclusions. In *Cochrane Handbook for Systematic Reviews of Interventions*. https://doi.org/10.1002/9781119536604.ch15

Schwab, J. F., & Lew-Williams, C. (2016). Language learning, socioeconomic status, and child-directed speech. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(4). https://doi.org/10.1002/wcs.1393

Segal, A., & Collin-Vézina, D. (2019). Impact of Adverse Childhood Experiences on Language Skills and Promising School Interventions. *Canadian Journal of School Psychology*, *34*(4). https://doi.org/10.1177/0829573519856818

Select Committee on Education and Skills. (2006). *Select Committee on Education and Skills: Third Report*. https://publications.parliament.uk/pa/cm200506/cmselect/cmeduski/478/47806.htm

Senthamarai Kannan, K., & Manoj, K. (2015). Outlier detection in multivariate data. *Applied Mathematical Sciences*, *9*(45–48). https://doi.org/10.12988/ams.2015.53213

Shepherd, P., & Gilbert, E. (2019). *Millennium Cohort Study: Ethical Review and Consent*. https://cls.ucl.ac.uk/wp-content/uploads/2017/07/MCS-Ethical-Approval-and-Consent-2019.pdf

Simoni, S. N. de, Leidow, I. C., Britz, D. L., Moraes, D. A. de O., & Keske-Soares, M. (2019). Impact of the speech sound disorders: family and child perception. *Revista CEFAC*, *21*(3). https://doi.org/10.1590/1982-0216/201921310718

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021). Framework for the development and evaluation of complex interventions: gap analysis, workshop and consultation-informed update. *Health Technology Assessment (Winchester, England)*, *25*(57). https://doi.org/10.3310/hta25570

Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. In *Educational Evaluation and Policy Analysis* (Vol. 31, Issue 4). https://doi.org/10.3102/0162373709352369

Smolak, E., McGregor, K. K., Arbisi-Kelm, T., & Eden, N. (2020). Sustained attention in developmental language disorder and its relation to working memory and language. *Journal of Speech, Language, and Hearing Research*, *63*(12). https://doi.org/10.1044/2020_JSLHR-20-00265

Snijders, V. E., Bogicevic, L., Verhoeven, M., & van Baar, A. L. (2020). Toddlers' language development: The gradual effect of gestational age, attention capacities, and maternal sensitivity. *International Journal of Environmental Research and Public Health*, *17*(21). https://doi.org/10.3390/ijerph17217926

Snow, K. L. (2006). Measuring school readiness: Conceptual and practical considerations. In *Early Education and Development* (Vol. 17, Issue 1). https://doi.org/10.1207/s15566935eed1701_2

Son, S. H., & Morrison, F. J. (2010). The nature and impact of changes in home learning environment on development of language and academic skills in preschool children. *Developmental Psychology*, *46*(5). https://doi.org/10.1037/a0020065

Song, F., Hooper, L., & Loke, Y. K. (2012). Publication bias: What is it? How do we measure it? How do we avoid it? In *Open Access Journal of Clinical Trials* (Vol. 5, Issue 1). https://doi.org/10.2147/OAJCT.S34419

Sørensen, L. B., Damsgaard, C. T., Petersen, R. A., Dalskov, S. M., Hjorth, M. F., Dyssegaard, C. B., Egelund, N., Tetens, I., Astrup, A., Lauritzen, L., & Michaelsen, K. F. (2016). Differences in the effects of school meals on children's cognitive performance according to gender, household education and baseline reading skills. *European Journal of Clinical Nutrition*, *70*(10). https://doi.org/10.1038/ejcn.2016.99

STATA. (2021). *vce options —Variance estimators*. https://www.stata.com/manuals/xtvce_options.pdf

Storkel, H. L., Voelmle, K., Fierro, V., Flake, K., Fleming, K. K., & Romine, R. S. (2017). Interactive book reading to accelerate word learning by kindergarten children with specific language impairment: Identifying an adequate intensity and variation in treatment response. *Language, Speech, and Hearing Services in Schools*, *48*(1). https://doi.org/10.1044/2016_LSHSS-16-0014

Su, X., Yan, X., & Tsai, C. L. (2012). Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(3). https://doi.org/10.1002/wics.1198

Sutcliffe, A. G., Barnes, J., Belsky, J., Gardiner, J., & Melhuish, E. (2012). The health and development of children born to older mothers in the United Kingdom: Observational study using longitudinal cohort data. *BMJ (Online)*, *345*(7876). https://doi.org/10.1136/bmj.e5116

Szpiro, A. A., Rice, K. M., & Lumley, T. (2010). Model-robust regression and a Bayesian “sandwich” estimator. *Annals of Applied Statistics*, *4*(4). https://doi.org/10.1214/10-AOAS362

Talbot, M. (2020). *Language and gender.* Polity Press.

Tawfik, G. M., Dila, K. A. S., Mohamed, M. Y. F., Tam, D. N. H., Kien, N. D., Ahmed, A. M., & Huy, N. T. (2019). A step by step guide for conducting a systematic review and meta-analysis with simulation data. In *Tropical Medicine and Health* (Vol. 47, Issue 1). https://doi.org/10.1186/s41182-019-0165-6

Taylor, C. L., Christensen, D., Lawrence, D., Mitrou, F., & Zubrick, S. R. (2013). Risk factors for children’s receptive vocabulary development from four to eight years in the longitudinal study of Australian children. *PloS One*, *8*(9). https://doi.org/10.1371/journal.pone.0073046

Tearne, J. E. (2015). Older maternal age and child behavioral and cognitive outcomes: A review of the literature. In *Fertility and Sterility* (Vol. 103, Issue 6). https://doi.org/10.1016/j.fertnstert.2015.04.027

Thabane, L., Mbuagbaw, L., Zhang, S., Samaan, Z., Marcucci, M., Ye, C., Thabane, M., Giangregorio, L., Dennis, B., Kosa, D., Debono, V. B., Dillenburg, R., Fruci, V., Bawor, M., Lee, J., Wells, G., & Goldsmith, C. H. (2013). A tutorial on sensitivity analyses in clinical trials: The what, why, when and how. In *BMC Medical Research Methodology* (Vol. 13, Issue 1). https://doi.org/10.1186/1471-2288-13-92

The Commonwealth Fund. (2020). *International health care system profiles: United States*. The Commonwealth Fund.

https://www.commonwealthfund.org/international-health-policy-center/countries/united-states

Thordardottir, E. (2010). Towards evidence-based practice in language intervention for bilingual children. *Journal of Communication Disorders*, *43*(6). https://doi.org/10.1016/j.jcomdis.2010.06.001

Thyer, B. A. (2012). *Quasi-experimental research designs.* . Oxford University Press.

Tierney, J. F., Stewart, L. A., & Clarke, M. (2021). Individual participant data. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions version* (6.2). Cochrane.

Tosh, R., Arnott, W., & Scarinci, N. (2017). Parent-implemented home therapy programmes for speech and language: a systematic review. In *International Journal of Language and Communication Disorders* (Vol. 52, Issue 3). https://doi.org/10.1111/1460-6984.12280

Trakulphadetkrai, N. V., Courtney, L., Clenton, J., Treffers-Daller, J., & Tsakalaki, A. (2020). The contribution of general language ability, reading comprehension and working memory to mathematics achievement among children with English as additional language (EAL): an exploratory study. *International Journal of Bilingual Education and Bilingualism*, *23*(4). https://doi.org/10.1080/13670050.2017.1373742

Tugwell, P., & Tovey, D. (2021). Should We Formalise the Assessment of Clinical Heterogeneity /Diversity In Systematic Reviews. In *Journal of Clinical Epidemiology* (Vol. 135). https://doi.org/10.1016/j.jclinepi.2021.05.015

Unterrainer, J. M., Ruh, N., Loosli, S. V., Heinze, K., Rahm, B., & Kaller, C. P. (2013). Planning steps forward in development: In girls earlier than in boys. *PLoS ONE*, *8*(11). https://doi.org/10.1371/journal.pone.0080772

Ursache, A., Blair, C., & Raver, C. C. (2012). The Promotion of Self-Regulation as a Means of Enhancing School Readiness and Early Achievement in Children at Risk for School Failure. In *Child Development Perspectives* (Vol. 6, Issue 2). https://doi.org/10.1111/j.1750-8606.2011.00209.x

Van Aert, R. C. M., Wicherts, J. M., & Van Assen, M. A. L. M. (2019). Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis. *PLoS ONE*, *14*(4). https://doi.org/10.1371/journal.pone.0215052

van den Bedem, N. P., Dockrell, J. E., van Alphen, P. M., de Rooij, M., Samson, A. C., Harjunen, E. L., & Rieffe, C. (2018). Depressive symptoms and emotion regulation strategies in children with and without developmental language disorder: a longitudinal study. *International Journal of Language and Communication Disorders*, *53*(6), 1110–1123. https://doi.org/10.1111/1460-6984.12423

Van Roy, B., Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *49*(12). https://doi.org/10.1111/j.1469-7610.2008.01942.x

Vauhkonen, T., Kallio, J., Kauppinen, T. M., & Erola, J. (2017). Intergenerational accumulation of social disadvantages across generations in young adulthood. *Research in Social Stratification and Mobility*, *48*, 42–52. https://doi.org/10.1016/j.rssm.2017.02.001

Verhoven, L., & Segers, E. (2003). Benefits of Speech Manipulation for Children With Language Disorders. In L. Verhoeven & H. van Balkom (Eds.), *Classification of Developmental Language Disorders: Theoretical Issues and Clinical Implications* (pp. 362–380). Psychology Press.

Vermeij, B. A. M., Wiefferink, C. H., Scholte, R. H. J., & Knoors, H. (2021). Language development and behaviour problems in toddlers indicated to have a developmental language disorder. *International Journal of Language and Communication Disorders*, *56*(6). https://doi.org/10.1111/1460-6984.12665

Vilma, A., & Johnson, J. (2020). *Millennium Cohort Study Data Handling Guide with syntax in R, STATA and SPSS*. UCL Centre for Longitudinal Studies. https://cls.ucl.ac.uk/wp-content/uploads/2020/09/MCS_Data_Handling_guide_ed1_2020-08-10.pdf

Volkers, N. (2018). Diverging Views on Language Disorders. *The ASHA Leader*, *23*(12). https://doi.org/10.1044/leader.ftr1.23122018.44

Vugteveen, J., de Bildt, A., Theunissen, M., Reijneveld, M., & Timmerman, M. (2021). Validity Aspects of the Strengths and Difficulties Questionnaire (SDQ)

Adolescent Self-Report and Parent-Report Versions Among Dutch Adolescents. *Assessment*, *28*(2). https://doi.org/10.1177/1073191119858416

Welsh, J. A., Bierman, K. L., & Mathis, E. T. (2014). Parenting programs that promote school readiness. In *Promoting School Readiness and Early Learning: The Implications of Developmental Reserach for Practice*.

Welsh, J. A., Bierman, K. L., Nix, R. L., & Heinrichs, B. N. (2020). Sustained effects of a school readiness intervention: 5th grade outcomes of the Head Start REDI program. *Early Childhood Research Quarterly*, *53*. https://doi.org/10.1016/j.ecresq.2020.03.009

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*, *91*(2). https://doi.org/10.1111/cdev.13219

Willinger, U., Deckert, M., Schmöger, M., Schaunig-Busch, I., Formann, A. K., & Auff, E. (2019). Developmental Steps in Metaphorical Language Abilities: The Influence of Age, Gender, Cognitive Flexibility, Information Processing Speed, and Analogical Reasoning. *Language and Speech*, *62*(2). https://doi.org/10.1177/0023830917746552

Willoughby, J. (2020). *The Economic Effects of Childhood Speech and Language Difficulties: A Cross-Cohort Quantitative Life Course Analysis*. Newcastle University.

Winsler, A., Hutchison, L. A., De Feyter, J. J., Manfra, L., Bleiker, C., Hartman, S. C., & Levitt, J. (2012). Child, family, and childcare predictors of delayed school entry and kindergarten retention among linguistically and ethnically diverse children. *Developmental Psychology*, *48*(5). https://doi.org/10.1037/a0026985

Winsler, A., Tran, H., Hartman, S. C., Madigan, A. L., Manfra, L., & Bleiker, C. (2008). School readiness gains made by ethnically diverse children in poverty attending center-based childcare and public school pre-kindergarten programs. *Early Childhood Research Quarterly*, *23*(3). https://doi.org/10.1016/j.ecresq.2008.02.003

Yew, S. G. K., & O'Kearney, R. (2013). Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments:

Meta-analyses of controlled prospective studies. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(5). https://doi.org/10.1111/jcpp.12009

Yim, D., & Yang, Y. (2018). The relationship of nonlinguistic visual processing capacity, speed, and vocabulary ability in children with specific language impairment. *Communication Sciences and Disorders*, *23*(2). https://doi.org/10.12963/csd.18501

Youth in Mind. (2016). Information for researchers and professionals about the Strengths & Difficulties Questionnaires. In *SDQ Info*. https://www.sdqinfo.org/

Zambrana, I. M., Ystrom, E., & Pons, F. (2012). Impact of gender, maternal education, and birth order on the development of language comprehension: A longitudinal study from 18 to 36 months of age. *Journal of Developmental and Behavioral Pediatrics*, *33*(2). https://doi.org/10.1097/DBP.0b013e31823d4f83

Ziv, Y. (2013). Social information processing patterns, social skills, and school readiness in preschool children. *Journal of Experimental Child Psychology*, *114*(2). https://doi.org/10.1016/j.jecp.2012.08.009