The economic cost of early vulnerable language skills

Madison E. Kerr, Jon Franklin

July 2021

PRO BONO ECONOMICS

Pro Bono Economics uses economics to empower the social sector and to increase wellbeing across the UK. We combine project work for individual charities and social enterprises with policy research that can drive systemic change. Working with 400 volunteer economists, we have supported over 500 charities since our inception in 2009.

National Literacy Trust is an independent charity which works with schools and communities to improve the literacy skills of disadvantaged children around the UK. We want to increase people's knowledge about the importance of early language skills and the tools available to them to improve the skills of children.

KPMG LLP, a UK limited liability partnership, operates from 21 offices across the UK with approximately 16,000 partners and staff. KPMG is a global network of professional firms providing Audit, Tax, Legal and Advisory services. For over 150 years, we've been supporting businesses to grow, our people to achieve and our communities to thrive.

Education Endowment Foundation is an independent charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

Foreword

Matt Whittaker, Chief Executive of Pro Bono Economics

Good language skills benefit us in so many ways – from our academic performance to our ability to build relationships, and from our sense of self-esteem to our chances of getting on at work. Indeed, as the fourth industrial revolution and the advancement of AI continue to change the nature of the labour market, language skills are likely to become an increasingly important driver of individuals' ability to find meaningful and rewarding work. In turn, they will be ever more crucial to economy-wide productivity. As such, they're worthy of significant investment.

And that investment needs to start early. The difficulties with language skills children display before entering formal education too often persist throughout their time at school, and beyond. As this report shows, our failure to create the right foundation for our pre-school children could be costing the country more than £300 million for every cohort that makes its way through education and employment.

This is certainly not a challenge that can be addressed by government and the education system alone. It is a challenge that requires government, businesses and parents to work alongside the passion, enthusiasm and creativity of social sector organisations such as the National Literacy Trust to deliver sustainable solutions that provide the environment, tools and support required to nurture our children's language skills.

This imperative has added urgency off the back of the Covid-19 crisis. The pandemic has almost certainly exacerbated the costs of not supporting early years language development, particularly for those from disadvantaged backgrounds. The need to address the challenges of early years language development is therefore pressing – both to prevent short-term impacts becoming long-term problems and to unleash the full potential of the next generation.

Bina Mehta, Chair of KPMG UK

The early years are a crucial stage in the education and development of children. Language skills developed in this period will impact a child's future experiences and opportunities. We know that one in four children who struggle with language at age five are more than twice as likely to be unemployed at age 34.

Prior to the pandemic, around 116,000 three-year-olds each year in the UK showed poor early language skills, leaving them likely to struggle with language development. COVID-19 has only worsened this, particularly for children from disadvantaged backgrounds.

Whilst the case for collaboration between business and schools is well-established, we must also turn our collective attention to the early years. Ensuring that the

youngest children in society can thrive is essential if we are to build a fairer, more prosperous and socially mobile society.

This report demonstrates that failure to improve early language skills could generate economic costs of £330m for each cohort of young people. Increasing levels of early language skills is vital to the sustainability and growth of UK business, both from an economic and employee pipeline perspective.

Support for the early years is integral to our long-term economic future, and that's why we have joined forces with the BBC's Tiny Happy People initiative and the National Literacy Trust to launch a business coalition dedicated to supporting early years language development. Together, we will play our role in increasing literacy, boosting the economy, and improving social mobility. We hope you will join us.

Jonathan Douglas CBE, Chief Executive of the National Literacy Trust

A child's early language and communication skills are not just the foundation of their literacy but influence a lifetime of social, emotional and economic outcomes. And, crucially, it is not just early years settings that create these skills but the stories and rhymes a child hears at home, the words games, the language they are exposed to in the context of the family that has a lifetime impact.

These early experiences, the home learning environment, are the root cause of much social inequity - middle and upper-income children are simply more likely to experience language rich early childhoods. But boosting the home learning environment offers protection against the effects of disadvantage for a child and is an important enabler of social mobility. The simple actions of sharing rhymes, stories and talking to babies can change society.

Never has this message been more important. The pandemic has exacerbated inequality and as recovery strategies are set the power of the home learning environment needs to be restated and placed centre stage.

Current pre-schoolers who are, or may become, at risk of having vulnerable language skills are facing a staggering lifetime cost of £1.2bn. By demonstrating this economic value, this report is a timely call to action to all who can support parents in the early years. Challenging not just the public and voluntary sectors but the business community who through products and services enrich so many early childhood experiences. As this report explains, more support for early parenting offers us a route to a more prosperous and fairer society.

Contents

Executive summary	6
Introduction	9
Background	11
Our approach	14
Results of our analysis	20
Conclusion and implications	25
Annexes	27

Executive summary

The development of early language skills provides a foundation for learning and is an important driver of later life outcomes. One in four children who struggle with language at age 5 do not reach the expected standard in English at the end of primary school. Children with a poor vocabulary at age five are more than twice as likely to be unemployed at age 34.

In this report, we have examined the long-term economic importance of early years language skills, highlighting why this basic skill is so crucial to individuals as well as wider society. We focus on a group of children that we describe as 'at-risk of vulnerable language skills'. Although there is no standard definition, we use this phrase to describe children who we believe may struggle with language skills in the future but are unlikely to suffer from a persistent language disorder. Results are calculated in 2020 prices with costs evaluated over the lifetime of children, up to the age of 60.

We find that:

- If the 14% of the current cohort of 3-year-olds in the UK at-risk of having vulnerable early language skills fail to improve their early language skills to above the at-risk threshold, it could generate lifetime economic costs of around £330 million.
- These costs accumulate across multiple cohorts of children – the lifetime cost of not supporting those in the four cohorts of pre-school children who are, or may become, at-risk of vulnerable language skills would amount to around £1.2 billion.
- This is the equivalent of £2,800 over the lifetime of each child at-risk of vulnerable language skills, with more than 95% of these costs likely to be in the form of decreased lifetime earnings for the children, with the remainder arising through increased costs for Special Educational Needs, involvement in the criminal justice system and demand on mental health services.
- However, research by Melhuish et al. (2017) has already demonstrated the positive impact that the home learning environment – the quality of activities a child is exposed to in the home that enable learning - can have on improving early years language skills and reducing these costs.
- If each of those 3-year-olds at-risk of vulnerable language skills did just one additional home learning activity per day such as reading, playing with letters or singing songs per day, it is estimated that it

could improve the early language skills of children, resulting in approximately £170 million in benefits for each cohort of 3-year-olds.

- If the average 3-year-old at-risk of vulnerable language skills did two additional home learning activities per day, estimates suggest it could be sufficient to lift them out of the at-risk of vulnerable language skills group.

Our study adds to the growing evidence of the importance of early language skills. Although there is significant uncertainty relating to some of the assumptions behind our analysis, sensitivity tests demonstrate that the broad conclusion – that there are substantial long-term costs from not supporting children who are at-risk of vulnerable language skills - remains robust to a wide range of plausible alternative assumptions.

There is a significant risk that these costs will have been further exacerbated as a result of the pandemic. The closure of many early years educational settings and the added stresses on parents and children mean the pandemic is likely to have negatively impacted the skill formation of pre-school children with evidence from Davies et al. (2021) suggesting that children from more disadvantaged backgrounds may have been disproportionately affected. It is imperative that these short-term impacts on development do not become long-term impacts on the academic and employment prospects of this generation's children. Recovery and development of early language skills can play a critical role in avoiding this.

With small but regular actions, like sharing books, singing or playing with letters more often, carers can have a transformational effect on the language development of very young children. Early language skills make a material difference both to individuals and the long-term prospects for the wider economy. Taking action to improve these important skills offers an opportunity to help our children's development recover following the crisis and address long-term structural inequality for disadvantaged families and areas.

14% of 3-year-olds in the UK are estimated to be at-risk of language vulnerabilities, this is the equivalent of around

116,000

children in the UK today

The potential lifetime cost from failing to support children at-risk of vulnerable language skills could be in the region of

£327m

for each cohort of 3-year-olds

These potential lifetime costs could reach

£1.2bn

for those in the four cohorts of pre-school children in the UK who are, or may become, at-risk of having vulnerable language skills

Doing one additional activity per day, such as reading, with 2-year-olds could reduce lifetime costs for those at-risk of early vulnerable language skills by

£166m

Introduction

The basic skills that children acquire during the first years of their lives are crucial. The development of social, emotional and cognitive skills by certain milestones provides children with the best chance of success when entering school. One of the most significant of these is children's ability to speak and to read – their language skills. One in four children who struggle with language at age five do not reach the expected standard in English at the end of primary school. Children with a poor vocabulary at age five are more than twice as likely to be unemployed at age 34.[1]

The pandemic has had dramatic impacts on the development of young people in the UK. During the first national lockdown in 2020, daily attendance at early years settings was less than a tenth of the level typically expected, and numbers remained low throughout the remainder of the year. They didn't fully recover to more typical levels until the spring 2021.[2]

There is an expectation that this will have serious effects on all children, but that those from disadvantaged backgrounds will experience them more acutely, widening already existing gaps in school readiness.[3] [4] Even outside formal settings, the added stress of pandemics on parents and children has been linked with developmental delay for children.[5] Early evidence suggests this is already having material impacts for schools. In a recent Education Endowment Foundation study, 96% of schools surveyed highlighted concerns with communication and language development of children in autumn 2020.[6]

This situation creates an urgent need to understand the long-term importance of early years language skills. Pro Bono Economics was commissioned by KPMG to provide an economic analysis of the long-term implications of early years language skills and explore what research can

---

[1] National Literacy Trust (2019): *Language unlocks reading; supporting early language and reading for every child,* National Literacy Trust.
[2] Data on attendance taken from Department for Education (2021): *Attendance in education and early years setting during the coronavirus(COVID-19) outbreak,* Table 2. Data on typical attendance taken from Department for Education (2020): *Statistics show children returning to early years settings,* Department for Education.
[3] Pascal C, Bertram T, Cullinan C, Holt-White E (2020): *COVID-19 and social mobility impact brief #4; early years,* Sutton Trust.
[4] Davies C, Hendry A, Gibson SP, Gliga T, McGillion M, Gonzalez-Gomez N (2021): *Early childhood education and care (ECEC) during COVID-19 boosts growth in language and executive function.* Infant and Child Development, e2241.
[5] de Araújo LA, Veloso CF, de Campos Souza M, de Azevedo JM, Tarro G (2020): *The potential impact of the COVID-19 pandemic on child growth and development: a systematic review.* Jornal de Pediatria.
[6] Bowyer-Crane C, Bonetti A, Compton S, Nielsen D, D'Apice K, Tracey L (2021): *The impact of Covid-19 on school starters; Interim briefing 1,* Education Endowment Foundation. Note that the sample was self-selecting so there is potential for bias in the results.

tell us about how these skills can be supported through activities in the home. This work has been completed in partnership with the National Literacy Trust to help evidence the importance of early years language skills and the role of parents, communities and businesses in this vital area of child development.

Scope of this study

Our study estimates the potential long-term economic costs of failing to improve the language skills of young children at-risk of having vulnerable language skills and assesses how changes in the home learning environment can help to address this.

There is no standard definition for children at-risk of vulnerable language skills. We therefore define this group as children who may struggle with language skills in the future but are unlikely to suffer from a persistent language disorder and would likely not require specialist support to achieve this improvement.[7]

We analyse data from the Department for Education's Study of Early Education and Development (SEED) to identify the gap in language skills for this group of children. To estimate the long-term economic costs of vulnerable language skills, we link this gap in language skills to the lifetime monetary value of not improving their early language skills.[8] We then draw on literature showing the positive relationship between the home learning environment and early language skills to discuss how improvements in the number of home learning activities can decrease these costs.

[7] The statistical definition of this group is discussed in detail in the methodology section of the report and Annex B.
[8] This report focuses on the economic costs of vulnerable early language skills; however, it does not estimate the costs associated with interventions to increase early language skills.

Background

The development of good communication skills early on in life is crucial for children's futures as not being able to communicate effectively can have significant effects on life opportunities. Evidence shows the importance of these skills before, during and after schooling.

Economic benefits of childhood language and literacy skills

There is extensive evidence of the link between children's language and literacy skills before school entry and their economic prospects later in life. In England, the SEED Value for Money report finds that a one standard deviation increase in naming vocabulary at age 3 is associated with £9,509 in potential lifetime long-term benefits per child, mostly attributed to increases in future employment and wages.[9] [10] Other benefits estimated include reduced costs related to Special Educational Needs, youth crime, smoking and depression. Similarly, evidence on improving language and reading skills for pre-school children in the USA estimates that a one standard deviation increase in kindergarten test scores at age 4 is likely to lead to an total increase in earnings up to age 66 between £15,000 and £26,000 per child.[11] An assessment of the Head Start programme estimated that a one standard deviation improvement in reading scores at age 3 and 4 is associated with £44,000 of increased lifetime earnings per child.[12] [13]

Literacy skills during school years have also been linked to substantial economic benefits.[14] [15] The Long-Term Cost of Literacy Difficulties study was carried out in 2009 by the Every Child a Chance Trust and the KPMG Foundation.[16] They estimated that neglecting to lift children out of literacy failure in primary school could lead to £6,400 to £82,000 in lifetime costs

[9] Paull & Xu (2017) uprated to 2020 prices.
[10] A standard deviation is a measure of the initial spread of observations for any outcome measure – it is often used in evaluations as a way of comparing changes between outcomes measured using different scales.
[11] Bartik TJ, Gormley W, Adelstein S (2012): *Earnings benefits of Tulsa's pre-K program for different income groups*, Economics of Education Review 31, no. 6: 1143-1161. Test scores are based on Woodcock-Johnson achievement tests covering Letter-Word ID, Spelling and applied problems. Estimated benefits are converted from US dollars to £ sterling and uprated to 2020 prices.
[12] See Duncan G, Ludwig J, Magnuson K (2010): *Child development,* In Targeting Investments in Children: Fighting Poverty When Resources are Limited, University of Chicago Press. They use the Peabody Picture Vocabulary Test designed to measure both receptive and expressive vocabulary. Estimated benefits are converted from US dollars to £ sterling and uprated to 2020 prices.
[13] Head Start is a programme funded by the United States federal government which provides low-income families with 0-5 year olds with courses and tools to improve the learning environment of children.
[14] Schoon I, Parsons S, Rush R, Law J (2010): *Childhood language skills and adult literacy: A 29-year follow-up study.* Pediatrics, *125*(3), e459-e466.
[15] National Literacy Trust (2019)
[16] Every Child a Chance Trust (2009): *The long-term cost of literacy difficulties.* This study is the second edition of a report commissioned by the KPMG Foundation in 2006.

per child.[17] A study by Pro Bono Economics and the KPMG Foundation examined evidence that improvements in the reading skills of 6 and 7-year-olds through the Reading Recovery programme had a range of fiscal benefits. These benefits included increased employment-related tax and national insurance revenues, reduced unemployment benefits, and savings on public services such as education, crime and healthcare over the life course of these children.[18] In total, the report found that improving literacy at age 6 through the Reading Recovery programme is associated with £9,900-£13,000 in lifetime benefits per child.[19]

Our study extends this previous research by estimating the scale of economic costs from not supporting a specific target group of children at-risk of having vulnerable language skills at age 3. Better understanding the long-term implications of poor language skills would have been important prior to the crisis but has particular value at a time when families, governments, businesses and communities are considering how to support children and young people to catch-up on the development opportunities lost during the pandemic.

Home learning environment

Children begin to learn long before entry into formal education. This makes the activities experienced by the child in the home vital to the development of their skills. The home learning environment characterises the quality of activities a child is exposed to that enable learning such as: sharing books, exposure to conversations, singing, playing with letters, etc.[20] Previous research has found that young children growing up in households with greater exposure to learning activities are associated with higher levels of cognitive and social development. [21] [22] [23] [24]

More specifically, studies have emphasised the importance of the home learning environment in terms of children's early language development.

---

[17] These estimates include additional outcomes that are not measured in our analysis such as: adult education costs, social costs of being NEET, health and crime costs. However, as in our study, the majority of costs arise in regards to employment outcomes. The wide range of costs is due to sensitivity analysis where they reduce the proportion of children that they estimate could be lifted out of literacy failure and also exclude health and crime costs. These estimates are uprated to 2020 prices.
[18] Franklin J, Kenward T, Pratt N (2020): *Assessing the impact of the Reading Recovery programme*. Pro Bono Economics.
[19] The Reading Recovery programme is intensive as it is daily on-to-one support from a trained specialist. These prices are uprated to 2020 prices.
[20] HM Government and National Literacy Trust (2018): Improving the home learning environment: A behaviour change approach.
[21] Melhuish E, Gardiner J, Morris, SP (2017): *Study of early education and development (SEED): Impact study on early education use and child outcomes up to age three*. Department for Education.
[22] Melhuish E, Gardiner J (2020): *Study of early education and development (SEED): Impact study on early education use and child outcomes up to age five years*. Department for Education
[23] Melhuish E (2010): *Impact of the home learning environment on child cognitive development: secondary analysis of data from "Growing Up in Scotland"*.
[24] These studies control for a large set of demographic variables.

The SEED study finds a positive association between the home learning environment at age 2 and naming vocabulary at age 3.[25] This study finds that the home learning environment has nearly the same effect on early language skills as maternal education, highlighting its importance in children's development. The impact of the home learning environment has also been shown to persist throughout childhood, through the positive association between improvements in the home learning environment on better academic outcomes at age 16.[26] [27]

National Literacy Trust advocates that initiatives which help families enrich the home learning environment can serve as a promising pathway to reduce structural inequalities.[28] More stimulating home learning environments are disproportionately found in households with higher levels of education and income.[29] Therefore, if government, employers and local communities can develop policies and tools to enable disadvatanged families to adopt these kinds of learning opportunities in their homes, it can support improved academic and employment outcomes that can contribute to breaking down intergenerational disadvantage.[30]

[25] Melhuish et al. (2017).
[26] Taggart B, Sylva K, Melhuish E, Sammons P, Siraj I (2015): *Effective pre-school, primary and secondary education project (EPPSE 3-16+): How pre-school influences children and young people's attainment and developmental outcomes over time.*
[27] Sammons P, Toth K, Sylva K, Melhuish E, Siraj I, Taggart B (2015): *The long-term role of the home learning environment in shaping students' academic attainment in secondary school.* Journal of Children's Services.
[28] https://literacytrust.org.uk/policy-and-campaigns/home-learning-environment/
[29] HM Government and National Literacy Trust (2018): *Improving the home learning environment: A behaviour change approach.*
[30] Serafino P, Tonkin R (2014): *Intergenrational transmission of disadvantage in the UK & EU,* Office for National Statistics

Our approach

Measurement of early years language skills

In this study we use the naming vocabulary measure from the British Ability Scales (BAS) III assessment as a measure of children's early language skills.[31] [32] This has been used in the Millennium Cohort Study and SEED studies as a measure of verbal abilities at early ages as it is available for children as young as 3 years old.[33] [34] [35] [36]

The economic costs of vulnerable language skills

In order to estimate the economic costs of vulnerable language skills, we take a four-step approach.

1

Define a threshold for children at-risk of having vulnerable language skills and calculate the improvement needed to raise the average child out of this category: Using the SEED dataset we define a threshold for children at-risk of having vulnerable language skills at age 3 as scoring one standard deviation or more below the mean naming vocabulary score.[37] Our analysis suggests that it would require 29% of a standard deviation increase in the naming vocabulary score to lift the average child out of the at-risk category.[38] Full details of our approach are available in Annex B.

[31] Elliot C, Smith P (2017): *British Ability Scales Third Edition* (BAS III). Windsor: GL-Assessment.
[32] We use the standardised age-adjusted naming vocabulary scores from the SEED dataset.
[33] Cullis A, Hansen K (2008): *Child development in the first three sweeps of the Millennium Cohort Study.*
[34] Paull G, Xu X (2017): *Study of Early Education and Development (SEED): The potential value for money of early education*, Department for Education Research report, Frontier Economics.
[35] More details on the naming vocabulary measure can be found in Annex B.
[36] This measure is also used in Law et al. (2012) as a measure of early language skills. Law J, Rush R, Anandan C, Cox M, Wood R (2012): *Predicting language change between 3 and 5 years and its implications for early identification.* Pediatrics, *130*(1), e132-e137.
[37] A standard deviation is a statistical measure that summarises the spread of data around its average. The threshold of one standard deviation below the mean is used in Melhuish et al. (2019) and Sammons et al. (2003) to define children at-risk of requiring Special Educational Needs at entry to pre-school. We have additionally assumed that those who score 1.5 or more standard deviations below the mean would require specialist support to improve their language skills based on research by Norbury et al. (2017). This coincides with 10% of the sample falling below this level which falls in line with current estimates for the proportion of children in the UK who have speech, language and communication needs (Norbury et al. 2016).
[38] Boys and children from lower income household are more likely to fall in this at-risk of vulnerable language skills category. Previous research using the Millennium Cohort Study also found that a mother with poor mental health, race and languages other than English being spoken in the household are associated with children being in the lowest quintile of naming vocabulary at age 3 (Cullis & Hansen 2008).

2

Use evidence from Paull & Xu (2017) to estimate long-term economic impacts of improving early language skills by one standard deviation: Paull & Xu (2017) use evidence from multiple sources to provide links from a one standard deviation increase in naming vocabulary at age 3 to later outcomes at age 10, 16 and 42 which can monetised such as: decreased probability of being in Special Education Needs, decreased probability of committing crime and increased probability of being employed and having higher wages. Further details about the Paull & Xu (2017) study are provided in Annex A.

3

Link the evidence to estimate the long-term economic impact of not raising the average child at-risk of vulnerable language skills to above the vulnerable threshold: we multiply the average improvement in the naming vocabulary measure necessary to raise the average child out of the vulnerable language category with the long-run monetary value of a one standard deviation improvement in naming vocabulary at the age of 3. The results of this analysis are provided in Figure 1.

4

Explore how improvements in early language skills can be achieved through improvements in the home learning environment: Using evidence from Melhuish et al. (2017) we explore two scenarios for how an increase in activities such as reading, playing with letters or singing songs at age 2 could improve children's early language skills at age 3 and reduce the long-term costs for those at-risk of vulnerable language skills:[39]

- Scenario 1: What impact would an additional home learning activity each day make to early years language skills?
- Scenario 2: What impact would two additional home learning activities each day make to early years language skills?

[39] Melhuish E, Gardiner J, Morris SP (2017): *Study of early education and development (SEED): Impact study on early education use and child outcomes up to age three*. Department for Education.

Throughout the report costs and benefits are expressed in 2020 prices and are discounted to age 3.[40] Figures are rounded to the nearest £1 for the purposes of transparency. There are inherent challenges in projecting lifetime outcomes for today's children as drivers of behaviour and earnings in the future could be affected by unforeseen structural changes in the economy over the intervening years. As such, our estimates should be viewed as indicative of the broad scale of potential costs.

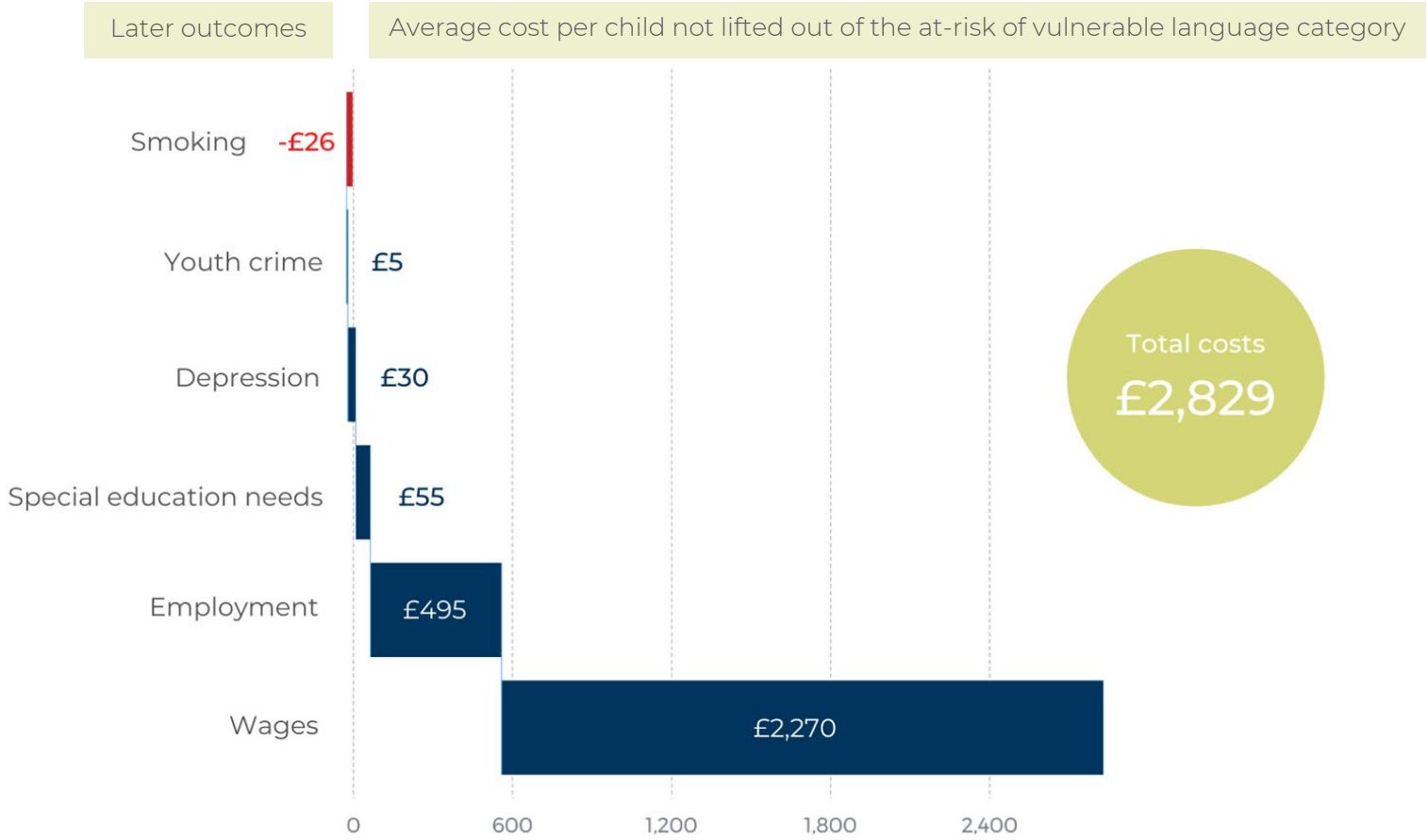
Figure 1 summarises our estimated costs associated with the at-risk of vulnerable language skills category. It shows that the average cost of not lifting a child at-risk of having vulnerable language skills out of this category is approximately £2,829 per child over their lifetime.

These costs are estimated by drawing upon Paull & Xu's (2017) research on the links between improvements in naming vocabulary at age 3 on later outcomes. They estimate the change in probability of these outcomes due to improvements in early language skills and provide monetary values for these impacts.

The estimated costs for failing to lift the average child out of the at-risk of vulnerable language skills category is accrued to both the individual child and the government over the child's lifetime up to age 60. The direct costs to the child include smoking, lost wages due to a reduction in the probability of being employed and lower wages if in employment.[41] The costs to the government include providing Special Education Needs, youth crime, health costs associated with smoking and higher benefit payments and lost taxation due to lower employment and wage prospects.

[40] Discounting involves the downward adjustment to flows of money in the future in order to reflect the preference of individuals and society to receive money sooner rather than later – it is standard practice for economic appraisals.
[41] Paull & Xu (2017) report a positive relationship between naming vocabulary and the probability of smoking resulting in this negative value for smoking.

Figure 1. Summary of economic impacts per child over their lifetime



Note: estimates have been discounted to age 3 and are shown in 2020 prices.

Key assumptions of the study

Our analysis is based on a number of assumptions, the most important of which are:

- The estimation of costs is based on a forecast of the expected impact of improvements in early language skills on later life outcomes, rather than on actual observed outcomes. Our analysis therefore relies on the assumption that the linkages between improved language skills in early years and later outcomes reported in Paull & Xu (2017) are both robust and relevant to children at-risk of having vulnerable language skills in the UK today. In particular, it should be noted that some of the longitudinal studies that underlie the Paull & Xu (2017) estimates use data that is now relatively old and measures that don't exactly align with the BAS measure used in the SEED study. However, there is some degree of reassurance from the fact that the monetary benefits from improved early years language skills in the Paull & Xu (2017) study are broadly in line with estimates using other methods for slightly later interventions in the UK and comparable studies in the USA.[42] Furthermore, we assume these estimated costs are constant over time meaning that the returns to improvements in early language skills have not changed.[43] The extent to which improvements in early language skills are sustained is uncertain but Cullis & Hansen (2008) find that children in the lowest quintile of early language skills at age 3 are significantly more likely to be found in the lowest quintile of language skills at age 5.
- We assume that the weighted sample of children in the SEED dataset are a nationally representative group of children in England; therefore, the threshold for those at-risk of vulnerable language skills and the subsequent percentage of children in the sample which fall below this threshold are assumed to be the same in the UK.
- The SEED dataset provides a measure of the home learning environment as an index. The index assumes that each activity is as important as another. In addition, there is no information on how long people do a given activity with a child, only the frequency of doing the activity. In reality, participating in home learning activities for longer periods could yield higher benefits for the child. We assume that families participating in these activities would have the

[42] See Bartik et al. (2012) and Duncan et al. (2010).
[43] It is possible that, as the UK economy has shifted more towards use of interpersonal skills, the wage loss from lower communication skills may have increased over time, suggesting that costs could be bigger. See for example, Dickerson A, Morris D (2019): *The changing demand for skills in the UK,* Centre for Vocational Educational Research Discussion Paper Series.

same average return to improvements in the home learning environment on naming vocabulary at age 3 as the families in the SEED dataset.

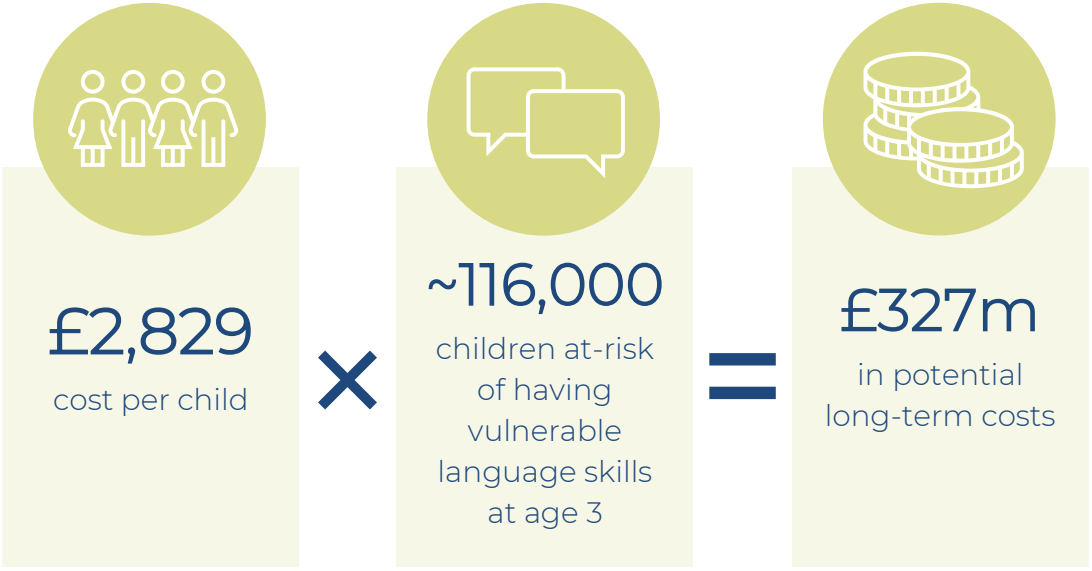
We explore the implications of some of these assumptions in a series of Sensitivity Tests in the following section. However, given the importance of these assumptions, we believe that our estimates of long-term costs should be treated as indicative of the broad scale of potential costs rather than as precise measures of long-term impacts.

Results of our analysis

Key findings

Overall, our core scenario suggests that not improving the early language skills of children at-risk of vulnerable language skills can potentially cost £2,829 per child over their lifetime. The majority of these costs come from the reduced probability of being employed and having lower wages, as shown in Figure 1 above. This cost per child is similar in scale to the cost of a disadvantaged child not attending an early education setting at age 2.[44]

Figure 2. Overview of costs



According to ONS estimates, there are approximately 800,000 3-year-olds in the UK.[45] If we assume that the UK population has the same proportion of 3-year-olds falling into the at-risk of having vulnerable language skills category as in England from the weighted SEED dataset, 14%, this would translate into approximately £327 million in long-term economic costs for each cohort of 3-year-olds that remain at-risk of vulnerable language skills. Further details on these calculations can be found in Annex B.

These costs accumulate across multiple cohorts of children – the lifetime cost of not supporting those in the four cohorts of pre-school children who

[44] See Table 22 of Paull G, Wilson C (2020): *Study of Early Education and Development (SEED): Financial returns to early education spending,* Department for Education
[45] ONS. (2019) *Population estimates for the UK, England and Wales, Scotland and Northern Ireland; mid-2019.*

are, or may become, at-risk of vulnerable language skills would amount to around £1.2 billion.[46]

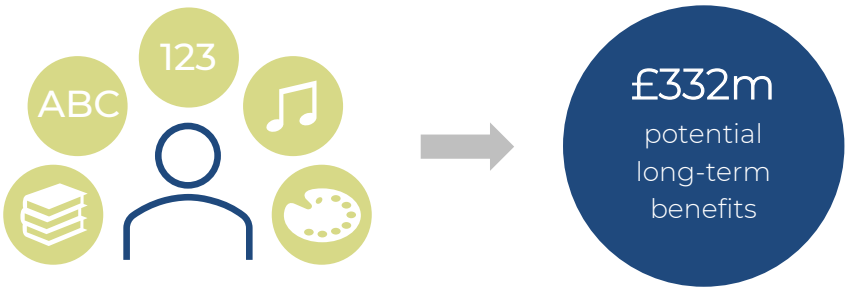
The impact of the home learning environment in addressing this cost

Drawing upon the evidence of the positive relationship between home learning environment at age 2 and naming vocabulary at age 3, we have examined the impact of two scenarios for improvements to the home learning environment in Figure 3.

Figure 3. Summary of home learning environment impacts

	Benefit per child	Total benefit
Scenario 1: One additional home-learning activity per day	£1,434	£166m
Scenario 2: Two additional home-learning activities per day	£2,868	£332m

If every child at-risk of having vulnerable language skills was supported in doing one additional home learning activity per day (reading, playing with letters, singing songs, etc.), this could translate to an estimated £166 million in potential long-term benefits.[47] Doing two additional activities per day is associated with an estimated £332 million in potential long-term benefits - sufficient to lift the average child who is at-risk of vulnerable language skills out of this category.[48] [49] Further details on calculations are provided in Annex C.



[46] This is based on assuming 14% of all children in the UK aged 3 and below are at risk of vulnerable language skills. We discount the costs for children younger than 3 to reflect the longer wait until these benefits are realised.
[47] These results assume that we can increase every child in the at-risk of vulnerable language skills home learning environment by those amounts.
[48] It takes approximately a 30% of a standard deviation increase in naming vocabulary to lift the average child out of the at-risk of vulnerable language skills category. Therefore, it takes a two standard deviation increase in the home learning index at age 2 to increase the BAS naming vocabulary by 30% of a standard deviation.
[49] These results assume that the child is not already doing the additional activities 7+ times a week.

Sensitivity analysis

To explore the robustness of our estimates of the cost of not improving the early language skills of children at-risk of vulnerable language skills we explore three alternative scenarios:

- **Sensitivity Test 1: Alternative assumptions for impacts of improvements in language skills on later outcomes:** In this analysis we explore the impact of altering the assumed long-term benefits from improvements in early year's language skills. To inform this we use the lowest and highest ends of the 95% confidence intervals estimated around the impacts on outcomes in the Carneiro et al. (2011) study that underlies the Paull & Xu (2017) estimates of economic benefits.[50] [51]
- **Sensitivity Test 2: Reduced lower threshold for individuals that need specialist support:** In our core analysis, we exclude children who are at-risk of having a persistent language disorder defined as those who score 1.5 or more standard deviations below the average naming vocabulary score as they may require additional support to improve their early language skills. In this analysis we explore the impact of assuming a smaller proportion of children would require specialist aid to increase their early language skills. This has the effect of increasing the proportion of the population that are in our target group of children who are at-risk of having vulnerable early language skills.
- **Sensitivity Test 3: Using reported national distributions for BAS scores rather than SEED data:** In this analysis we explore the impact of calculating the threshold for children at-risk of vulnerable language skills using the distribution of the naming vocabulary measure that was used when creating the original standardised measure, rather than relying on the observed scores in the SEED dataset.

Sensitivity Test 1: Alternative assumptions for impacts of improvements in language skills on later outcomes

To account for the difficulty of predicting associations between outcomes separated by decades (in this case, language skills at age 3 and, for example, future earnings of those children in later life), we run the analysis using the upper and lower 95% confidence intervals for estimates of the associations between cognitive skills at age 7 and later outcomes as found

[50] A 95% confidence interval is a way of representing the uncertainty around a central estimate – it shows a range within which we can be 95% sure that the true value lies.
[51] Carneiro P, Crawford C, Goodman A (2011): *The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes*. University College London, Institute for Fiscal Studies and Georgetown University, American Economic Association 2012 Annual Meeting Conference Paper

in Carneiro et al. (2011) and used in Paull & Xu (2017).[52] Further details about Carneiro et al. (2011) can be found in Annex A. If we adopt the lower end of the 95% confidence intervals for estimated effect sizes in Carneiro et al. (2011), then this results in a reduced cost per child over their lifetime of £2,224 which translates to a total cost of £257 million. If we adopt the higher end of the 95% confidence intervals, this results in an increased net benefit per child over their lifetime of £3,444 which translates to a total cost of £398 million.

Sensitivity Test 2: Lower threshold for children that may need specialist support

In our core scenario, we assume that a portion of children are at-risk of having a persistent language disorder and may require specialist support to improve their early language skills. We have excluded these children for the purposes of our analysis as the type of interventions to support them is likely to be different from that required for our 'at-risk of vulnerable language skills' category. In the core analysis, we refer to Norbury et al. (2017) and define the group of children who may have language disorders as 1.5 standard deviations and below the average naming vocabulary score. This corresponds with 10% of the sample falling in this category.[53] [54]

In this sensitivity test, we assume that a smaller percentage of children would fall into this lower category, in line with evidence suggesting that the proportion of children with speech, language and communication needs could be as low as 7%.[55] [56] Assuming that the same relationship exists between improvements in early language skills and later life outcomes, using this reduced lower threshold results in an estimated lifetime cost of £3,371 per child and total costs of £471 million.

Sensitivity Test 3: Using reported national distributions for BAS scores rather than SEED data

When creating the British Ability Scales III naming vocabulary measure, GL Assessment used a nationally representative sample of children to create a standardised distribution of the measure. The naming vocabulary measure

---

[52] Paull & Xu (2017) link naming vocabulary at age 3 to these age 7 cognitive measures.
[53] Norbury CF, 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.
[54] The 10% consists of 7% of the lowest scoring children in the naming vocabulary assessment and 3% from the SEED dataset who were not able to take the naming vocabulary assessment due to speech, hearing or other difficulties.
[55] Law J, Charlton J, Dockrell J, Gascoigne M, McKean C, Theakston A (2017): *Early Language Development: Needs, provision, and intervention for preschool children from socio-economically disadvantage backgrounds*. London: Education Endowment Foundation, 204.
[56] House of Commons Library (2018): *Speech, language and communication support for children.*

from the SEED dataset that we use in the analysis in this report is calculated in relation to the national distribution.

This sensitivity test examines the impact on the results if we define our threshold for children at-risk of vulnerable language skills using the national distribution mean (score=50) and standard deviation (10).[57] This means that children scoring 40 and below are categorised as at-risk of vulnerable language skills instead of a score of 39 and below in the core scenario using the SEED dataset.[58] If we calculate our threshold based on this distribution, this results in an estimated lifetime cost of £3,092 per child and total costs of £442 million.

Conclusions of sensitivity analysis

The results of our sensitivity analysis are summarised in Figure 4 below. We find that our alternative assumptions do have an impact of the level of costs indicating the sensitivity of the results to the assumptions made. However, in each case, the results demonstrate that failing to increase the early language skills of all children that are at-risk of language vulnerabilities is associated with significant long-term potential costs, estimated to be in the region of £300 - £500 million.

Figure 4. Summary of sensitivity analysis

	Cost per child	Total Cost[59]
Core Scenario	£2,829	£327m
Sensitivity Test 1a: Upper 95% confidence interval	£3,444	£398m
Sensitivity Test 1b: Lower 95% confidence interval	£2,224	£257m
Sensitivity Test 2: Lower threshold of 7%	£3,371	£471m
Sensitivity Test 3: National distribution	£3,092	£442m

[57] This is not dissimilar from our sample's mean of 51.5 and standard deviation of 11.8.
[58] In this scenario we keep the same lower threshold at a score of 33 and below to keep in line with the 10% of the population falling in the category of at-risk of having a persistent language disorder.
[59] The total cost accounts for how many children in the UK would fall into the at-risk of vulnerable language skills but do not require specialist assistance category. In Sensitivity Test 2, more children fall into the at-risk of vulnerable language skills category as we assume fewer children would require specialist assistance. In Sensitivity Test 3, the score which indicates vulnerable language skills is one point higher than with our sample, therefore, there are more children in the at-risk of vulnerable language skills category resulting in a larger total cost than in Sensitivity Test 1.

Conclusions and implications

At a time when the development of early years language skills of our children has been detrimentally affected by the pandemic, our study has demonstrated that this could have meaningful long-term consequences for children over their lifetimes and the economy more broadly.

Our analysis suggests that:

- If the 14% of the current cohort of 3-year-olds in the UK at-risk of having vulnerable early language skills fail to improve their early language skills to above the at-risk threshold, it could generate lifetime economic costs of around £330 million.
- These costs accumulate across multiple cohorts of children – the lifetime cost of not supporting those in the four cohorts of pre-school children who are, or may become, at-risk of vulnerable language skills would amount to around £1.2bn.
- This is the equivalent of £2,800 over the lifetime of each child at-risk of vulnerable language skills, with more than 95% of these costs likely to be in the form of decreased lifetime earnings for the children, with the remainder arising through increased costs for Special Educational Needs, involvement in the criminal justice system and demand on mental health services.
- However, research by Melhuish et al. (2017) has already demonstrated the positive impact that the home learning environment – the quality of activities a child is exposed to in the home that enable learning - can have on improving early years language skills and reducing these costs.
- If each of those 3-year-olds at-risk of vulnerable language skills did just one additional home learning activity per day such as reading, playing with letters or singing songs per day, it is estimated that it could improve the early language skills of children, resulting in approximately £170 million in benefits for each cohort of 3-year-olds.
- If the average 3-year-old at-risk of vulnerable language skills did two additional home learning activities per day, estimates suggest it could be sufficient to lift them out of the at-risk of vulnerable language skills group.

Implications

Our study adds to the growing evidence of the importance of early language skills by demonstrating the potential long-run costs of not helping children at-risk of vulnerable language skills. The conclusions have been tested under a range of alternative assumptions; and whilst the exact estimate of costs varies, the conclusion that there are large economic costs for failure to increase the early language skills of children at-risk of having vulnerable language skills remains robust.

There is a significant risk that these costs will have been further exacerbated as a result of the pandemic, however, evidence suggests that steps can be taken to address this. There is good evidence to show that parents and carers can make a real difference in tackling these costs through improvements in the home learning environment.[60] Creating opportunities for children to engage in practical home learning activities such as reading, playing with letters and singing songs could go a long way to closing the gap in language skills for those at-risk of having vulnerable early language skills.

Previous research has demonstrated significant variations in childhood attainment across regions in the UK, with 68% of children achieving a "good level of development" at age 5 in the South East compared to just 55% in the East Midlands.[61] Geographic differences in childhood attainment can be partially explained by differences in socio-economic characteristics across regions such as levels of employment and educational level of mothers.[62] Based on the evidence in our study, improving support for early years language skills for the most disadvantaged areas would be expected to improve the long-term employment prospects for children in these areas, helping reduce structural inequalities found in society.

[60] Melhuish et al. (2017).
[61] Dunatchik A, Wishart R, Cartagena-Farias J, Smith N (2018): Regional differences in attainment in the early years.
[62] Dunatchik et al. (2018).

Annex A – Details on Paull & Xu (2017)

In this section we provide further details of the evidence we have used to quantify the monetary value of improvements in early language skills. We review this in two parts:

- Part 1: Links from BAS naming vocabulary at age 3 to later outcomes in a child's life
- Part 2: Estimates of the monetary values of improvements in these later outcomes associated with improvements in the BAS naming vocabulary at age 3

Part 1: Links from BAS naming vocabulary at age 3 to later outcomes in a child's life

Paull & Xu (2017) draw on a retrospective study by Carneiro et al. (2011) to establish the link between BAS naming vocabulary in early years and several outcomes in later life. The Carneiro study is based on longitudinal data from the National Child Development Survey (NCDS), which is a cohort study based on all individuals born in Great Britain in a single week in March 1958. It uses this data to assesses the relationship between children's cognitive skills at age 7 and the likelihood of:[63]

- youth crime and smoking by age 16; and
- depression, employment and hourly wages at age 42.

There are some limitations of using this evidence:

- First, our use of the Carneiro et al. (2011) study means we assume that the links between BAS naming vocabulary and cognitive measures in early years and later outcomes for children today are similar with the NCDS sample as several decades ago.
- Second, the NCDS does not include the same cognitive measures as research Paull & Xu (2017) use to draw links between measures at different ages. Therefore, Paull & Xu (2017) assume that cognitive skills measured at the same age have one-to-one associations with each other. For example, Paull & Xu assume that a one standard deviation increase in BAS naming vocabulary at age 5 is equivalent to a one standard deviation increase in the Millennium Cohort Study

[63] Paull & Xu (2017) draw links between naming vocabulary at age 3 and children's cognitive skills at age 7.

cognitive measure at age 5.[64] Furthermore, Paull & Xu assume the relationships found between Carneiro et al. (2011) with the NCDS cognitive score measured at age 7 on later outcomes to be the same as the MCS cognitive score measured at age 7 on later outcomes. In practice the correlations between these measures are unlikely to be one-to-one.

Paull & Xu (2017) link improvements with the BAS naming vocabulary to requiring Special Educational Needs through results in a study by Anders et al. (2011).[65]

The links between early development and later outcomes that are relevant to our study are shown in Figure 5 below. This figure shows the change in the probability of each later outcome that is associated with a one standard deviation improvement in the BAS naming vocabulary score at age 3. These figures are based on the associations shown in Table 5 of Paull & Xu (2017) which are then converted to associations based on a one standard deviation increase in BAS naming vocabulary at age 3 as shown in Figure 2 in Paull & Xu (2017). Paull & Xu state that these associations are statistically significant and based on regression analysis with extensive controls for other factors.

Part 2: Estimates of the monetary values of improvements in these later outcomes associated with improvements in the BAS naming vocabulary at age 3

Paull & Xu (2017) calculate the estimated lifetime value of the outcomes listed above from the Carneiro et al. (2011) and Anders et al. (2011) studies to enable them to estimate the economic impact of improvements in the BAS naming vocabulary measure at age 3. These benefits arise over an individual's life between the ages of 5 and 60, depending on the outcome. Paull & Xu draw on several sources of evidence to estimate the relevant monetary values (see Table 6 of Paull & Xu for details) which are selected to avoid double-counting of benefits.

We use the monetary values associated with improvements in later outcomes relating to a one standard deviation increase in BAS naming vocabulary in early years that are reported in column 1 of Table 7 of Paull & Xu (2017). These values represent the expected change in the lifetime value

[64] The MCS cognitive measure includes the BAS naming vocabulary measure along with two other measures of cognitive skills.
[65] Anders Y, Sammons P, Taggart B, Sylva K, Melhuish E, Siraj-Blatchford I (2011): *The influence of child, family, home factors and pre-school education on the identification of Special Educational Needs at age 10.* British Educational Research Journal, *37*(3), 421-441.

per child associated with the change in probabilities for each of the outcomes reported in Figure 5 below.[66]

There are some limitations of using this evidence:

- To estimate the lifetime value of the outcomes, Paull & Xu (2017) assume that the relationships found in Carneiro et al. (2011) of increases or decreases in probability on later outcomes associated with an increase of cognitive score at age 7, are the same for other ages where the benefit is applicable. For example, they assume that a one standard deviation increase in BAS naming vocabulary at age 3 is associated with 2.1% higher wages at each age between 16 and 60.
- Furthermore, we use updated estimates of the costs of crime from Paull & Xu (2017) to reflect the unit costs of crime information from the Home Office.[67]

[66] The values in Table 7 of Paull & Xu relate to the present value of the future benefits from improved outcomes associated with a one standard deviation increase in BAS naming vocabulary at age 3, discounted to age 3.
[67] Home Office (2018).

Figure 5. Summary of Paull & Xu (2017) impacts of improvements in naming vocabulary

Later outcome	Impact of a one standard deviation improvement in BAS naming vocabulary at age 3 on outcome in Paull & Xu Figure 2	Lifetime value of outcome found in Paull & Xu Table 6	Reported monetary value of one standard deviation increase in BAS naming vocabulary at age 3 in Paull & Xu Table 7
Special Educational Needs	-0.36% SEN reading -0.06% SEN maths	-£44,072	£186
Youth crime	-0.18%	-£7,855	£14
Smoking	0.15%	-£58,781	-£87
Depression	-0.35%	-£29,060	£101
Employment	0.39%	£425,996	£1,664
Wages	2.1%	£360,826	£7,633

Note: estimates have been discounted to age 3 and are shown in 2020 prices. Results may not align due to rounding.

Annex B – Details of cost calculations

In this section we provide further details for our approaches to assessing the costs of not improving the early language skills of children for children at-risk of having vulnerable language skills.

- **Step 1: Define threshold for at-risk of vulnerable language skills:** providing further information on how we calculate this threshold and the improvement in naming vocabulary score necessary to raise the average child out of the at-risk of vulnerable language skills category.
- **Step 2: Use evidence from Paull & Xu (2017) to estimate long-term economic impacts of improving early language skills by one standard deviation:** further details on how our cost estimates were derived from the Paull & Xu (2017) findings.
- **Step 3: Link the evidence to estimate the long-term economic impact of improving early language skills for children at-risk of vulnerable language skills**: we multiply the percent of a standard deviation improvement in naming vocabulary necessary to lift the average child out of the at-risk of vulnerable language skills category (from Step 1) with the long-run monetary impacts of a one standard deviation improvement in naming vocabulary (from Step 2) to provide an assessment of the long-run average cost of not improving the early language skills of children at-risk of vulnerable language skills.

Data: Study of Early Education and Development (SEED)

This report uses the SEED dataset to calculate the threshold for at-risk of vulnerable language skills group of children. This dataset is based on a large-scale longitudinal study following the progress of a cohort of approximately 4,500 children in England from age 2 through age 4. The SEED study was carried out by NatCEN Social Research, working with Frontier Economics, the University of Oxford and Action for Children. The baseline survey ran in 2013-2014 with subsequent waves measured in 2014-2015 and 2015-2016. In our analysis, we used a final sample of 4,326 children and adopted the inverse-probability weights to account for different selection probability and bias due to non-response, making the sample representative of families with two-year-olds in England.[68]

[68] Speight S (2015): *Study of early education and development: baseline survey of families*. Research report, July 2015.

Measure of early language skills: British Ability Scales Naming Vocabulary

In this study we use the naming vocabulary measure from the British Ability Scales (BAS) III assessment as a measure of children's early language skills.[69] [70] This measure has been used in the Millennium Cohort Study, SEED, EPPE and EPPSE studies as a measure of verbal abilities at early ages as it has been validated for children as young as 3-years-old. More specifically, it is a measure of expressive language and knowledge of object names.[71] [72] The measure is constructed by showing children a series of pictures of objects and are asked to say the name of the object.

Step 1: Define threshold for children at-risk of having vulnerable language skills

We break this part of the analysis into four key steps:

a) Define threshold for those at-risk of having vulnerable language skills

b) Define lower threshold for those who are at-risk of having a persistent language disorder and may require specialist support to improve early language skills

c) Calculate the average improvement necessary to lift the average child out of the at-risk of vulnerable language skills category.

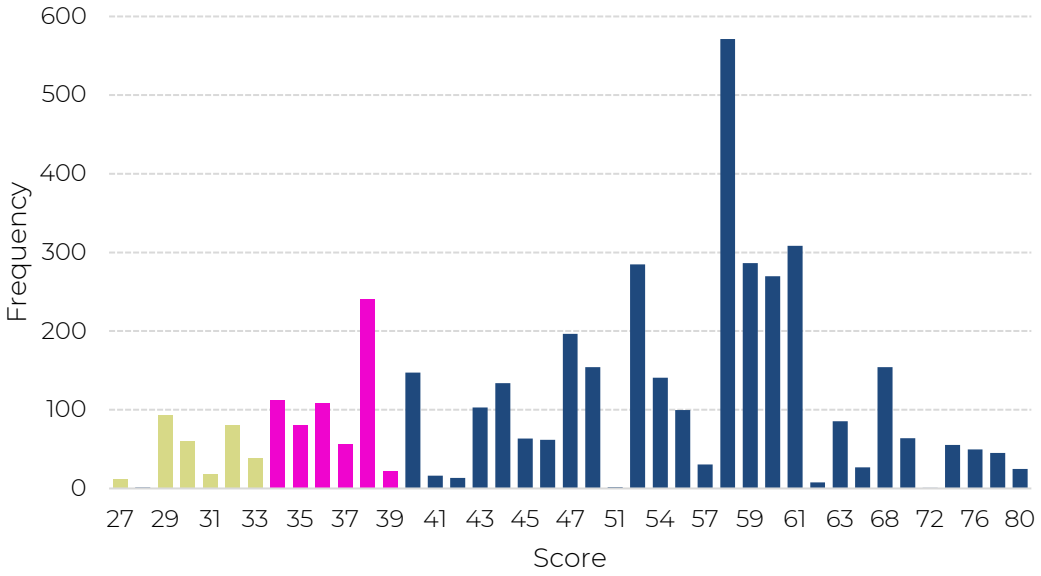
d) Convert the average improvement necessary to lift a child out of the at-risk category to percent of a standard deviation increase so it can be linked with Paull & Xu (2017) estimates in Step 3

a) There is no formal definition of a child at-risk of having vulnerable language skills; therefore, we draw upon previously published research in the field of child development to inform our choice of threshold. Previous research on estimating the risk of children developing Special Educational Needs has used the threshold of one standard deviation or more below the mean of a cognitive score for children at entry to pre-school.[73] [74] We therefore, define the group of children who are one standard deviation or more below the mean naming vocabulary score to be at-risk of having vulnerable language skills. With a mean of 51.5

[69] Elliot & Smith (2017).
[70] We use the standardised age-adjusted naming vocabulary scores from the SEED dataset.
[71] Elliot & Smith (2017).
[72] As this measure is on expressive language, children who are reluctant to speak may receive low scores.
[73] Melhuish E, Barnes J, Gardiner J, Siraj I, Sammons P, Sylva K, Taggart B (2019): *A study of the long-term influence of early childhood education and care (ECEC) on the risk for developing special educational needs (SEN).* Exceptionality Education International. 29(3), 22-41.
[74] Sammons P, Taggart B, Smees R, Sylva K, Melhuish E, Siraj-Blatchford I, Elliot K (2003): *The Early Years Transition & Special Educational Needs (EYTSEN) Project.*

and standard deviation of 11.8, this coincides with a threshold for children at-risk of having vulnerable language skills as those scoring 39 and below on the naming vocabulary assessment, as shown in Figure 6.

Figure 6. Distribution of naming vocabulary at age 3 in SEED dataset



b) The scope of this research is to estimate the costs of not improving the early language skills of children at-risk of having vulnerable language skills. We assume that a portion of children are at-risk of having a peristent language disorder and may require specialist support to improve early language skills. For the purpose of our study, these children are not included in the at-risk of vulnerable language skills category as the interventions and support required to address their needs is likely to be different.[75] Drawing upon research by Norbury et al. (2017), we categorise children who score 1.5 standard deviations or more below the mean as children who may require additional investment in their language skills.[76] This results in a lower threshold of a score of 33 and below as shown in green in Figure 6. This coincides with 10% of the sample falling below this level which falls in line with current estimates

[75] By excluding the children at the lowest end of the distribution from our target group of children, and our cost estimates, we provide a conservative estimate of the costs associated with failing to improve the early language skills of children.

[76] Norbury CF, 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.

for the proportion of children in the UK who have speech, language and communication needs.[77] [78]

c) To estimate the average improvement in naming vocabulary necessary to lift the average child out of the at-risk of vulnerable language skills category, we calculate the average naming vocabulary score for the children in the at-risk category but above the lower threshold (as shown in pink in Figure 6). This approach is summarised in the formula below:

$$Improv = Vocab_{Not\ vulnerable} - \overline{Vocab}_{vulnerable}$$

Where:

$Improv$ = average improvement in BAS naming vocabulary score to lift child out of at-risk of vulnerable language skills category (3.52 points)
$Vocab_{Not\ vulnerable}$ = BAS naming vocabulary score outside of the at-risk of vulnerable language skills (score=40)
$\overline{Vocab}_{vulnerable}$ = average BAS naming vocabulary score of children at-risk of having vulnerable language skills and above the lower threshold of children who may require additional assistance to increase their language skills (score=36.48)

In our sample from the SEED dataset, the average child who is at-risk of having vulnerable language skills and can potentially have their early language skills improved is 36.48. This means that on average it would take an improvement of 3.52 points in the BAS naming vocabulary score to lift the average child above our threshold of 39 and out of the group at-risk of having vulnerable language skills.

d) To link the improvement in the BAS naming vocabulary necessary to lift the average child out of the at-risk of having vulnerable language skills category to Paull & Xu's (2017) estimates for the economic value of increases in BAS naming vocabulary, we must calculate what percent of a standard deviation this 3.52 points is equivalent to. To do this, we use the formula below:

$$Improv_{Deviation} = \frac{Improv}{\sigma_{Full\ Sample}}$$

[77] Norbury CF, Gooch D, Wray C, Baird G, Charman T, Simonoff E, Vamvakas G, Pickles A (2016): *The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study.* Journal of child psychology and psychiatry, *57*(11), 1247-1257.

[78] The 10% in this lower threshold consists of 7% in the sample who took the naming vocabulary assessment and scored 33 or below and 3% from the SEED dataset who were not able to take the naming vocabulary assessment duenot being able to speak, hear or another reason This estimate coincides with a population study by Norbury et al. (2016) which estimates that approximately 2.34% of children experience "language impairment associated with intellectual disability and/or existing medical diagnosis" It is important to incorporate these children into the sample as when the results are extrapolated to the UK, we must take into account that there is a portion of children who are unable to take the assessment.

Where:

$Improv_{Deviation}$ = the percent of a standard deviation improvement in BAS naming vocabulary score to lift child out of at-risk of vulnerable language skills category (29.73%)

$\sigma_{Full\ Sample}$ = standard deviation of the BAS naming vocabulary measure for the full sample in the SEED dataset (11.83)

In our sample from the SEED dataset, it requires 3.52 points to lift the average child out of the at-risk of vulnerable language skills category and with a standard deviation of the full sample BAS naming vocabulary measure of 11.83, this results in a 29.73% of a standard deviation increase in the BAS vocabulary score to lift the average child out of the at-risk of vulnerable language skills category.

Steps 2 and 3: Use evidence from Paull & Xu (2017) to estimate long-term economic impacts of improving early language skills by one standard deviation

We use several key pieces of information from Paull & Xu (2017) to estimate the long-term economic value from a one standard deviation improvement in BAS naming vocabulary at age 3. The costs are already discounted to age 3 but we inflated prices to be more up to date:

- We take the unit costs for the different outcomes from Table 7 of Paull & Xu (2017) and inflate them from 2015/16 prices (used in Paull & Xu) to 2020 prices using ONS's GDP deflator.
- We then multiply this by the percent of a standard deviation required to lift the average child out of the at-risk of vulnerable language skills category found above

This approach is summarised in the formula below:

$$B_j = \sum_{j=1}^{7} b_j * \pi * Improv_{Deviation}$$

Where:

B_j = average cost savings per child from a 0.29 standard deviation increase in BAS naming vocabulary, for outcome j, discounted to age 3 in 2020 prices

b_j = unit cost savings per child for an improvement in outcome j associated with a one standard deviation increase in BAS naming vocabulary, discounted to age 3

π = factor for price differences between 2016 and 2020 based on ONS GDP deflator, equal to 1.13

Figure 7 shows that the average cost of not lifting a child at-risk of having vulnerable language skills out of this category is approximately £2,829 per child over their lifetime.

Figure 7. Impacts of improvements in naming vocabulary at age 3

Later outcome	Reported monetary value of one standard deviation increase in BAS naming vocabulary at age 3 in Paull & Xu Table 7 (b_j)	Reported monetary value of one standard deviation increase in BAS naming vocabulary at age 3 in 2020 prices ($b_j * \pi$)	Average cost per child not lifted out of the at-risk of vulnerable language category (B_j)
Special Educational Needs	£165	£186	£55
Youth crime	£14	£16	£5
Smoking	-£77	-£87	-£26
Depression	£90	£101	£30
Employment	£1,478	£1,664	£495
Wages	£6,778	£7,633	£2,270
Total costs	**£8,448**	**£9,513**	**£2,829**

Notes: the impact of a one standard deviation improvement in naming vocabulary on unit outcome is the increase or decrease in probability of that outcome for all benefits excluding increased wages. Increased wages are measured as the percent increase in wages on average. Further details on total costs of outcomes calculated in Paull & Xu (2017) can be found in Annex A.

Extrapolate results to the UK

According to ONS estimates, there are approximately 800,000 3-year-olds in the UK.[79] Assuming that the UK has the same proportion of children falling into the at-risk of vulnerable language skills as estimated by the SEED dataset for England, enables us to calculate the potential economic cost of not lifting 3-year-olds in the UK out of the at-risk of vulnerable language skills using the following formula:

$$Costs_{Total} = B_j * [N * (Prop_{Vulnerable} - Prop_{Lower})]$$

Where:

$Costs_{Total}$ = total lifetime costs of not improving early language skills of 3-year-olds at-risk of having vulnerable language skills (£327m)

N= estimate of number of 3 year olds in the UK by the ONS (800,000)

$Prop_{Vulnerable}$ = the proportion of children that are at-risk of having vulnerable language skills and below (0.2446)

$Prop_{Lower}$ = the proportion of children that are at-risk of having a persistent language disorder and may require additional inputs to improve early language skills (0.10)

The lifetime individual cost of not improving early language skills extrapolated to the UK population translates to approximately £327 million in long-term economic costs of not improving the language skills of the average 3-year-old child at-risk of having vulnerable language skills to a minimum standard.

[79] ONS. (2019) *Population estimates for the UK, England and Wales, Scotland and Northern Ireland; mid-2019.*

Annex C – Details of home learning environment

In this section we provide further details about the evidence relating to the home learning environment at age 2 and how it is associated with early language skills at age 3.

Previous research has shown the importance in the home learning environment for the development of children's early language development.[80] [81] In this section we describe how families can potentially increase the early language skills of children through different activities in the home which promote learning.

In the SEED dataset, the children's home learning environment is assessed through the construction of a home learning index based on how many times a week someone in the family does the following with the child:

1. Reads
2. Plays with letters
3. Plays with numbers
4. Teaches songs/rhymes/poems
5. Paints or draw

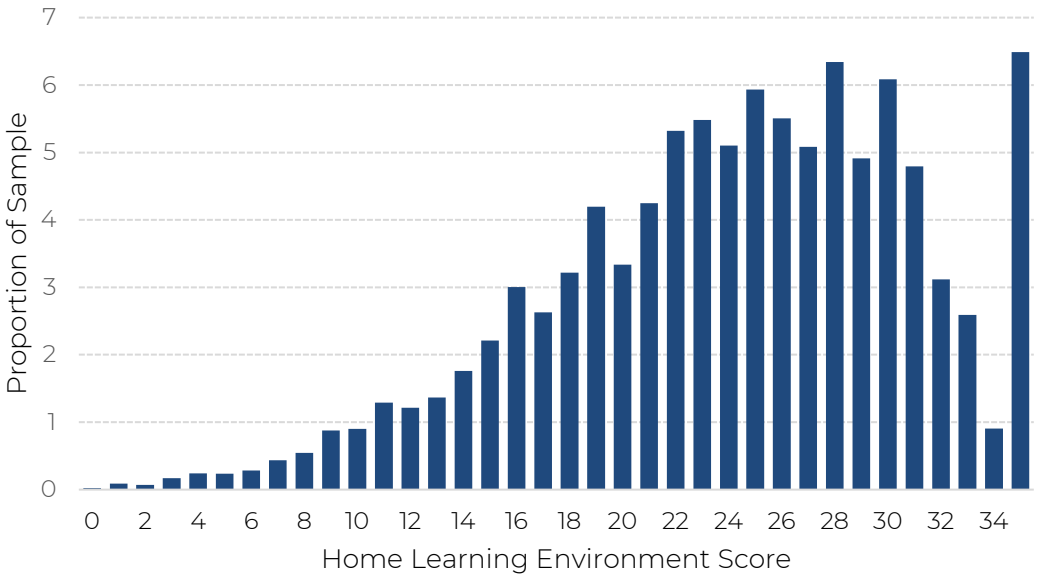
They state how often a week someone does each of the activities with a child through the following responses:[82]

- Never
- Occasionally (less than once a week)
- 1 or 2 times a week
- 3 times a week
- 4 times a week
- 5 times a week
- 6 times a week
- 7+ times a week

[80] Melhuish et al. (2017).
[81] Melhuish & Gardiner (2020).
[82] There is not data available as to how long each activity is performed.

The home learning index is comprised of the summation of how often someone does the above five activities with the child a week with a score range of 0-35.[83] The distribution of the home learning index from the SEED dataset is shown in Figure 8. The average child in the sample did approximately 24 activities a week. The standard deviation of the sample is equal to 6.85 (which is approximately 7 activities a week).

Figure 8. Distribution of home learning environment at age 2 in SEED dataset



Previous research shows that a one standard deviation increase in the home learning environment score at age 2 is associated with a 15% of a standard deviation increase in BAS naming vocabulary at age 3 while controlling for a host of demographic variables.[84] Following the same analytical method as above, we can estimate the potential long-term benefits of improvement in the home learning environment due to the links between the home learning environment and the BAS naming vocabulary measure. First, we calculate the percent of a standard deviation increase in the home learning environment of a given scenario with the following formula:

[83] The construction of the index assumes that each activity is as important as another as it assume that an increase in the weekly frequency a child is read to is equivalent to the same increase in the weekly frequency of playing with letters.
[84] Melhuish et al. (2017).

$$HLE\_Improv_{Deviation} = \frac{HLE\_Improv}{\sigma_{HLE\ Full\ Sample}}$$

Where:

$HLE\_Improv_{Deviation}$ = the percent of a standard deviation improvement in the home learning environment index

HLE_Improv = the increase in number of weekly activities done with the child (i.e. one or two more activities a day)

$\sigma_{HLE\ Full\ Sample}$ = standard deviation of the home learning environment index for the full sample in the SEED dataset (6.85)

To calculate the benefits associated with increases in the home learning environment, we use the following formula:

$$B_{HLE} = \sum_{j=1}^{7} b_j * \pi * HLE\_Improv_{Deviation}$$

Where:

B_{HLE} = benefit per child from an increase in the home learning environment, discounted to age 3 in 2020 prices

Then to extrapolate the individual benefits to those of all 3-year-olds at-risk of having vulnerable language skills we use the following formula:

$$Benefits_{HLE\ Total} = B_{HLE} * [N * (Prop_{Vulnerable} - Prop_{Lower})]$$