Best of both worlds? Early cognitive and non-cognitive development of children from native and mixed families.

Joanna Clifton-Sprigg*

Draft - provisional and incomplete

June 11, 2014

Abstract

This paper looks at the early childhood development of children with at least one foreign parent. Using the data for families with children under the age of 6 in Scotland, I find that *mixed* families are better off on average than natives but families with two foreign parents are worse off. These groups also differ in terms of activities they engage in and views related to children's upbringing. Mixed family and native children perform differently in exercises related to linguistic ability, but their overall cognitive and non-cognitive skills are similar. Regression results suggest that children with at least one foreign parent score 6% less than native children in the Vocabulary Naming Exercise, but the gap narrows with age. However, the difference for children whose both parents are foreign-born reaches almost 21%. Some investments, including practicing letters and visits to zoo or museum, have higher returns in terms of cognitive development for mixed families. The closing of the gap with age is consistent with current theories about development of bilingual children. The results, although potentially subject to bias, are informative and provide new information about early childhood development. I consider a growing in size, but so far under-researched, family type and compare performance of children at an early age. I shed light onto the mechanisms behind the initial performance gap and its closing with age and demonstrate that there is an additional disadvantage in terms of language-related performance to having two, rather than one foreign-born parent.

JEL-Classification: I25, J13, J15, F22
Keywords: Childhood, Children, Education, Immigration

*Advice given by Prof. Maia Güell, Prof. Mike Elsby and Dr Steven Dieterle has been a great help in the execution of this project. I would also like to thank those involved in the Growing Up in Scotland project for data access as well as valuable advice on how to use it. My research work has also been supported by the ESRC scholarship.

1 Introduction

Lanzieri (2012) reports that, between 2006 and 2007, 15% of all marriages registered in Europe were mixed marriages between natives and foreigners. He attributes the growing phenomenon to large scale migratory movements in the European Union. Mixed families are faced with different difficulties relative to native families regarding their children's upbringing. Many of them function in two languages and combine two or more cultures. Hence, several interesting questions arise; how do children from such households function in the society and, in particular, how do they perform in terms of cognitive and non-cognitive skills relative to their native peers, given the environment they have been exposed to? What environment do parents in *mixed* families create for their children? Are children from *mixed* families disadvantaged and should governments be proposing policies to target the group?

This matter is interesting for economists concerned with human capital accumulation at the early stages in life (Heckman and Conti, 2012) and with emergence of performance gaps in children. Educational inequalities emerge early and should be targeted as soon as they appear to prevent social inequality (Nores and Barnett (2010), Carneiro, Heckman, et al. (2011)). Moreover, human capital is key for economic growth (Hanushek and Woessmann, 2009).

In this paper I compare the socio-economic characteristics of mixed and native families in Scotland who have children under the age of 6. I look at their lifestyle as defined by the activities they engage in and their children's performance reflected in various measures of cognitive and non-cognitive skills and physical development.

I am predominantly interested in the children's performance. They are often growing up in a challenging and stimulating bilingual environment, adapting to two, or more, cultures. From a linguistic point of view, they may experience a delay in speech due to bilingualism (Baker, 1999), which may temporarily affect their cognitive and non-cognitive skills, although linguists argue that they remain largely unaffected over the long term (Kaushanskaya and Marian, 2007).

Children's performance will be substantially shaped by their parents' actions. Although *mixed* families do not face the same assimilation challenges as fully *foreign* families, they nonetheless may lack location-specific knowledge essential for a child's upbringing, e.g. institutional arrangements, welfare entitlements, etc. They may also have smaller networks than *native* families. These factors may obstruct a child's development by limiting its exposure to and interaction with *native* children. On the other hand, foreigners entering intermarriage are usually positively selected in terms of education and socio-economic background (Lanzieri, 2012). They are in a better position to create a favourable upbringing for children and consciously engage in activities which enhance children's cognitive and non-cognitive development.

How these factors balance out in this situation is the empirical question I would like to shed light on in this paper. I exploit a rich *Growing Up in Scotland* (GUS) dataset created by the Scottish Government which provides information about a large, randomly selected sample of children born in Scotland: their family situation, socio-economic characteristics, every day activities they engage in, parents' opinions on various matters related to their children as well as a series of tests of cognitive and non-cognitive performance of the children. I use a *mixed* family status, i.e. whether child has at least one parent born outside of the UK, along with the language spoken at home as indicators of the bicultural and bilingual environment.

I find that *mixed* and *native* families differ in socio-economics, with higher percentage of *mixed* households falling into the top socio-economic categories, higher percentage of respondents in *mixed* families having a university degree and creating a full, rather than one parent, family. Despite being seemingly positively selected, the families match the portrait of immigrant residents of Scotland based on various data sources. Although they engage in a similar amount and type of physical activity, *mixed* and *native* families differ in the way they spend time with their children. Respondents from *mixed* families more often practice letters with

their children, take them to library or museum. Children in *mixed* families also use computer less frequently and watch less TV on average. Parents differ in views and ambitions too. *Mixed* family parents are more likely to hope for their child to go to university and less likely to see single parenthood as better than having both parents. These outcomes may reflect' differing lifestyles and family situation.

Children perform comparably in most fields, with an exception of a Vocabulary Naming exercise and a Symbolic Composite of the Communication and Symbolic Behaviour score, in which *mixed* family children perform worse, both in conditional and unconditional comparisons. These measures are heavily dependent on the linguistic advancement of a child. Regressions with an interaction term between *mixed* family status and *bilingual* status confirm that language is a significant contributor to the difference in performance. The initial gap narrows with age, but does not close by the age of 5, the last point at which children were assessed. There is no differential impact by gender. I analyse whether the contribution of various activities to the Vocabulary Naming score differs across the groups and find that some, but not all factors, including practising letters with a child, visits to the zoo, museum and swimming pool, have a higher payoff for *mixed* family and bilingual children. This is expected if worse performance is related to linguistic skills as these activities facilitate language acquisition.

It is important to note at the outset that mixed families are not equivalent with the families where both parents are foreign. *Fully foreign* families constitute a very small proportion of the GUS sample (1%) and have been included in the definition of *mixed family group* as a comprehensive analysis of outcomes for such a small subset is not possible. However, to point out the disparity between the two groups, I compare the family characteristics and run an introductory conditional analysis of outcomes, separating the *mixed* and *fully foreign* families. Whilst *mixed* families are positively selected relative to *natives*, the *fully foreign* families are more disadvantaged. A higher percentage of *fully foreign* families have a degree qualification, relative to natives, but they are more likely to be classified in lower occupational categories, have lower incomes and to live in the most deprived areas of Scotland. Children from *fully foreign* families underperform in the same aspects of cognitive and non-cognitive development as children from *mixed* families, but the gap between them and *native* children is much larger. For example, in the Vocabulary Naming Exercise, children from mixed families score about 4.5% and from fully foreign families almost 21% lower than native children. This may be because children from *fully foreign* families not only face a linguistic obstacle, but their parents may be in a worse position to help them catch up due to lack of the institutional, cultural and linguistic knowledge necessary to do so.

The analysis is rather descriptive in nature and relies on raw comparisons and OLS regressions, reliability of which is threatened by unobserved heterogeneity and selection bias. I discuss the extent to which these factors are problematic and underline the fact that results I obtain are robust to inclusion of a variety of controls. Despite the shortcomings, this paper confirms many findings from the literature to date and further contributes to it on various fronts.

Firstly, I consider *mixed* families whose situation has not been researched in depth so far. The literature focused on second generation immigrants or, more specifically, second generation minority immigrants. However, the 'in-between' group exposed to both native and foreign culture due to the intermarriage is growing and may be getting the best (or the worst) of both worlds. It also differs significantly from *fully foreign* families. Thus, it is important to understand how the upbringing process differs in such environments and whether families need to be supported by policy.

Another important novelty of this paper is the analysis of performance of children under the age of 6. The majority of literature has examined children from ages 5 and above, through high school and often into adulthood. It could not provide the answer to how early, if ever, the gap emerges. Studies which looked at small children (Reardon and Galindo (2009), Fuller,

Bridges, et al. (2009)) focus on the US and its Latino communities, which are a very specific example of this situation and compare fully *foreign* families with *natives*.

Further, I have at my disposal various, often complementary, measures of development, which allows me to conclude that cognition, broadly speaking, is not affected by the *mixed* family status and that the only effects are related to linguistic skills.

The paper is structured in the following way. I provide a brief overview of relevant literature in Section 2. Then I discuss data and provide unconditional comparisons in Section 3. Section 4 contains regression analysis and its results. In Section 5 I discuss limitations of my approach and conclude.

2 Literature

This paper reflects ideas from various strands of economic, sociological and linguistic literature. Economists have long argued that development of cognitive and non-cognitive skills is vital for short-term (Apps, Mendolia, et al. (2012)) and long-term outcomes of individuals (Aizer and Cunha (2012), Behrman, Hoddinott, et al. (2006), Feinstein (2003)) and plays an important role in economic development (Hanushek and Woessmann, 2009).

These skills develop very early in life (Heckman and Conti (2012), Carneiro, Crawford, et al. (2007)) and depend on the initial level of human capital as well as investments made, which are complementary. What kinds of investments are most effective has been subject to a debate (Keane and Fiorini, 2012). Moreover, performance gaps between children emerge early and are best tackled as soon as they appear (Heckman and Conti (2012), Nores and Barnett (2010)) as early investments yield high returns to society (Conti, Heckman, et al. (2011), Cunha, Heckman, et al. (2010)).

Although policy interventions are key for eliminating social inequality, parental roles in early childhood are indispensable in bridging the gap. Human and cultural capital are transmitted across generations and can influence educational outcomes (Black, Devereux, et al. (2005), Black and Devereux (2011), Holmlund, Lindahl, et al. (2011), Bjorklund and Salvanes (2010)). Children's attitudes towards school, aspirations and non-cognitive skills are highly correlated with those of their parents (Heckman and Rubinstein (2001), Borghans, Lee Duckworth, et al. (2008), Carneiro, Crawford, et al. (2007)). Activities families engage in and lifestyle, which form cultural capital, are usually learnt from parents and have influence on cognitive and non-cognitive skills (Meier Jaeger, 2011). De Philippis (2014) argues that culture is so persistent, it can explain correlation in PISA test scores between second generation immigrants and natives in their home countries.

Research suggests that family characteristics such as income and education (Ermisch (2008), Hartas (2011)) but also time spent reading, writing or practising rhymes (Melhuish, Sylva, et al., 2008) all determine children's cognitive and non-cognitive performance. Keane and Fiorini (2012) find that time spent in educational activities is the most productive input into cognitive skill development. Thus, it is crucial to account for these elements in the analysis.

The considerations just discussed also play a role in the context of migration. Economic literature established the existence of a performance gap for first generation immigrants, relative to the native population. The extent of the difference and whether it disappears with time depend crucially on the age at arrival in the country (Boehlmark, 2008) as well as the length of stay before the gap is measured (Glick and Hohmann-Marriott (2007), Glick, Hanish, et al. (2012)). The divide is also visible for second generation immigrants but varies across countries (Dustmann, Frattini, et al., 2012). In fact, studying second generation immigrants from minority groups in Britain, Dustmann, Frattini, et al. (2010) find that, for some minorities, the pupils not only catch up but even outperform their native peers. For this group whether the gap

closes depends, among other factors, on ethnicity and country of birth (Reardon and Galindo (2009), Glick, Hanish, et al. (2012)), parental education levels (Fuller, Bridges, et al., 2009) and language spoken at home (Dustmann, Frattini, et al. (2010), Rosenthal, Baker, et al. (1983)). Activities parents engage in are also central to the discussion (Brooks-Gunn and Markman, 2005). For example, Becker (2010) finds that Turkish children benefit more from activities outside the household than their German peers in terms of language development.

Most studies focus on immigrants past the early childhood stage. Dustmann, Machin, et al. (2010) consider 5-16 year olds in the UK, whilst Dustmann, Frattini, et al. (2012), Dronkers and de Heus (2012) and Kornder and Dronkers (2012) look at 15 year olds in Europe and Nordin and Rooth (2007) look at labour market outcomes of grown up second generation immigrants. With the exception of a few studies, little is known about immigrant children's performance at earlier stages of life[1]. Reardon and Galindo (2009) look at development of cognitive and non-cognitive skills of pre-schoolers and Fuller, Bridges, et al. (2009) of toddlers, but they focus specifically on Latino communities in the US. Hence, my analysis adds to the work in this area.

Further, research has generally focused on second generation immigrants and the literature on performance of children from mixed marriages is rather limited. Duncan and Trejo (2011) study outcomes of 16-17 year olds from Mexican-American mixed families and find that they outperform other Mexican second generation immigrants. They do not compare the group with the native population though.

In the paper I argue that language, specifically simultaneous bilingualism, may be the main channel of difference in performance between children, which brings me to linguistics literature. Baker (1999) provides an extensive overview of the impacts of bilingualism on cognitive outcomes in children. Bilinguals seem to have an advantage in certain thinking dimensions, such as divergent thinking, creativity, early metalinguistic awareness and communicative sensitivity. At the same time, bilingual children may initially possess a smaller vocabulary in each of their languages (Oller and Eilers (2002), Portocarrero, Burright, et al. (2007), Bialystok (2009)). Nonetheless, so far research found no correlations between bilingualism and IQ (Kaushanskaya and Marian, 2007) and it is suggested that many cognitive skills remain unaffected by bilingualism (Baker, 1999). Most recent research indicates that bilingualism can slow down cognitive ageing by exerting a positive effect on later-life cognition (Bak, Nissan, et al., 2014).

3 Data and descriptives

3.1 Data

The data used in this analysis come from the Growing Up in Scotland (GUS) longitudinal study gathering information about physical, cognitive and non-cognitive development of children born in Scotland, as well as demographic and socio-economic details of the households they live in. The main topics covered by the study include the household composition and family background (parental education, income, employment, etc.), parental relationships, support parents receive and their views on parenting, childcare, pre-school and subsequently school enrollment, the child's health and development, the activities the child is involved in (incl. outings, physical and intellectual activities at home), social networks and children's development assessments. The participating families were randomly selected using Child Benefit records for Scotland and data was further weighted to adjust for initial selection as well as attrition. I apply the longitudinal weights throughout the analysis. For more details of the selection and weighting procedures, see Appendix A.

The study now captures three cohorts of children: Birth cohort 1 (BC1) of circa 5000

[1]Note that the list proposed here is exemplary and by no means exhaustive.

children born in 2004/2005, Birth Cohort 2 (BC2) of about 6000 children born in 2010/2011 as well as Child Cohort (CC) of around 3000 born in 2002/2003. Due to data availability I rely on the BC1 and CC data for the purpose of this analysis[2]. The data for CC comprise 4 annual waves following children from age 3 to 6; the data for BC1 has been collected for 6 annual waves from when the children were 10 months of age until 6 years old. I apply the relevant weights and then combine the data to focus on analysis by age, rather than cohort. At a final wave the achieved sample size for both cohorts is 5857, with almost 8% of children from *mixed families*. This group participated in all waves of the study but observations are also available for those who participated only in some waves.

3.2 Identification of children from mixed families

Given that almost all children in the sample were born in Scotland, I am interested in *second generation immigrants*. In particular, I define a child as coming from a *mixed family* if at least one of its parents was born outside of the UK and as *native* if both parents were born in the UK. For the most part of the analysis children from the *fully foreign* families are included in the group of *mixed family* children[3]. However, I also repeat the introductory regressions separating the two groups; in that one case the *mixed family* status indicates that one parent is foreign and one native and *fully foreign* that both parents were born outside of the UK.

Definition of migrant status on the basis of country of birth is standard in the literature (Ozden, Parsons, et al., 2011) but has its limitations, as I cannot distinguish certain groups from each other. For example, a parent born abroad to two British citizens who then moved back to the UK will be identified as foreign born in this study. Equally a parent who is a second generation immigrant himself will be identified as *native* as he was born in the UK. In the majority of such cases I expect, however, the definition to imply that a child is brought up by parents of different nationalities, cultures and potentially in two different languages. Migration status is also often determined on basis of one's nationality, but this too has its drawbacks and is impossible to apply in this case, as no nationality information was collected during GUS.

Another way of identifying the children from *mixed* families is on the basis of language spoken at home. Using the data at hand, I separate groups who speak only English at home, English and another language and another language only (negligible group). This measure is less accurate and is likely to further underestimate the number of *mixed* families, as many foreign born residents of Scotland may be native speakers of English (e.g. if they come from the USA or Australia, etc.) or may choose to speak English, rather than their first language at home. Moreover, it is possible that speakers of Gaelic or Scots identified themselves as speaking "other language" at home even though they are native residents of Scotland. I have no way of estimating the size of this group.

Although intertwined, language and migration status may have different implications for children's development. Admittedly, there may be heterogeneity in development of children from *mixed* families depending on the language spoken at home, which is an additional factor in development of cognitive and non-cognitive skills. At early stages of development, children who speak the native language of the country they are growing up in, may find it easier to assimilate and interact with society (Rosenthal, Baker, et al., 1983).

These two defining variables are closely related, with the correlation coefficient of .594. As can be seen in Table 1, 49% of children in *mixed families* speak English and another language

[2]So far only one wave of data for Birth Cohort 2 has been released and it is not as informative for the purpose of this analysis. More information about the study and resultant research can be found on the project website, growingupinscotland.org.uk

[3]I included children with two foreign parents and with one foreign and one native parent in the same group because the former group constitutes a very small proportion of the sample. There were only 70 children with both foreign parents in the combined sample in the final wave.

or another language only at home. The corresponding group among *natives* reaches only 1%. I will be using both of them independently of each other to see whether the results depend on the definition. From here on *mixed* family status is defined by the variable *mixed* and *bilingualism* is identified by the variable *language*. It will become clear that often they are equivalent in terms of the results I obtain.

Table 1: Sample size and language

Panel A: Sample size at final wave			
	mixed families	fully foreign families	native families
Birth cohort	318	45	3344
Child cohort	179	25	2021
Total	497	70	5365
Panel B: Language spoken at home			
	only English	English and other	other only
overall	94%	5%	1%
mixed and foreign family	51%	40%	9%
native family	99%	1%	0%
Panel C: correlation - language and mixed			0.594

3.3 Outcome variables for children

I analyse various measures of cognitive and non-cognitive development which were collected for participating children. As a check and to argue that there are unlikely to be differences in other aspects of development, I also briefly look at measures of motor and physical development available in the data. Below I describe how the outcome variables were created as well as which cohort and age group they are available for.

The Strength and Difficulties Questionnaire (SDQ) is a behavioural screening questionnaire and was undertaken for children in both cohorts at ages 4, 5 and 6. The scale includes 25 questions used to measure five aspects of a child's development - emotional symptoms, conduct problems, hyperactivity or inattention, peer relationship problems and pro-social behaviour. A score is calculated for each aspect and the total score is a sum of the scores from all the scales except the pro-social. The main indicator, total SDQ score, is a variable on the scale of 0-34 with the higher score indicating a worse performance.

The Communication and Symbolic Behaviour Score (CSBS) measures non-cognitive development of children and was only used with Birth Cohort children at the age of 2. Respondents had to complete questions which assessed their child's communication, emotional development, understanding and interaction with peers. The 24 questions were grouped into clusters of individual scores. Clusters can be added into three composite scores assessing social communication, expressive language and symbolic functioning. A total score is the sum of the three composites and ranges from 0 to 57, with the higher score indicating better performance.

The exact questions and groupings which contribute to each score in SDQ and CSBS can be found in Appendix A.

British Ability Scales are measures of cognitive development and contain two exercises: Naming Vocabulary and Picture Similarities tests. The vocabulary test is aimed at assessment of spoken vocabulary and involves the child naming coloured pictures from a booklet he is shown. The exercise captures expressive language ability as well as the recall skill and depends on the child's existing vocabulary. The Picture Similarities test consists of a booklet with four

images on each page and a set of cards with a single image. The child is asked to match the card with a picture in the booklet on the basis of them sharing an element or a concept. I use the percentile normative scores in analysis. The normative scores are derived from standard tables and defined with the reference to the standardisation sample used in developing of the assessment (see: Bradshaw, Tipping, et al. (2009) for details).

Respondents were also asked to assess **the child's speech development** from age 2 onwards. This is a subjective measure which was based on whether: 1) the child can be understood by strangers, 2) the child can be understood by family and friends and 3) the child can be understood by the respondent. The answer was to be given on the scale from 1 to 3 where 1 indicated mostly, 2 sometimes and 3 not at all.

Children in both cohorts were also assessed in terms of their **physical and motor development**. The test for babies took place at the age of 1 and for toddlers at the age of 3. Hence, CC was tested only once (age 3) and BC was subject to a baby test at age 1 and to a toddler test at age 3.

Availability of the outcomes for both cohorts at any given age is presented in Table 2.

Table 2: Availability of outcome measures across cohorts and age

Age	1		2		3		4		5		6	
	BC1	CC	BC1	CC	BC1	CC	BC1	CC	BC1	CC	BC1	CC
SDQ score							×	×	×	×	×	×
CSBS score			×									
BAS score					×				×			
Child's speech			×		×	×	×	×	×			
Motor development	×				×	×						

Note: here × indicates that data are available for this age group and cohort.

3.4 What do we learn about Scottish families - unconditional analysis

In this section I investigate differences between families in socio-economics, the way they spend time, their views and attitudes and their children's performance. I use weighted data, but do not control for any other characteristics in the comparisons.

3.4.1 Household composition and socio-economic situation

The *mixed* and *native* families in the study differ from each other in socio-economic characteristics. The families are similar in size, but a higher percentage of respondents in *native* families are lone parents in comparison with *mixed families*. In particular, 17% of native parents were lone parents when their child was 6 in contrast to only 11% of parents from *mixed* families.

Pronounced differences emerge also in terms of education with 47% of mixed parents and 29% of native parents having completed a degree. The higher educational attainment in *mixed* families is only partly channelled into their equivalised household incomes which are comparable with those of natives, except for the bottom quintiles. *Mixed* households are more likely to be classified higher in the NS-SEC classification with 67% falling into managerial and professional classification, compared with 53% of the native households. A higher percentage of *mixed*

families live in the 20% least deprived areas of Scotland. Thus, so far I find no indication of children in *mixed* families being at any material disadvantage relative to native children. In fact, given higher educational attainment of their parents on average, one may be inclined to conclude the opposite.

On the other hand, *fully foreign* families appear to be disadvantaged relative to *native* families. Even though almost 40% of parents from *foreign* families have a degree qualification, only 42% are employed in managerial and professional occupations. Notably, a higher percentage of them are small business owners when compared to *mixed* and *native* families. These families are also overrepresented in the bottom quintile of the household income distribution and almost a third live in the 20% most deprived areas of Scotland.

For most of the analysis I group *mixed* and foreign families together. Summary statistics for the group can be found in the first column of Table 3 and closely reflect the outcomes of the *mixed* families as they dominate in terms of size.

The observations about socio-economic characteristics prompt the question of potential selection and representativeness of the group classified as *mixed* families. The concern is justified by the evidence in literature that more educated immigrants have a higher propensity to intermarry with natives (Sandefur and McKinnell (1986), Lichter and Qian (2001), Meng and Gregory (2005), Chiswick and Houseworth (2011)) and the theory of assortative mating (Greenwood, Guner, et al., 2014), suggesting that the group may be positively selected.

Table 3: Summary statistics - family socio-economics

% of lone parents	**ALL foreign**	***mixed only***	***fully foreign***	**Native**
Child's age	(N=567)	*(N=70)*	*(N=497)*	(N=5365)
3	4.61	5.2	1.6	16.89
4	5.97	7	0	15.71
5	7.98	9.17	1.27	17.31
6	9.76	11.1	1.43	17.05
Household NSSEC category (%) ***	**ALL foreign**	***mixed only***	***fully foreign***	**Native**
Managerial and professional occupations	63.37	*67.4*	*41.9*	52.62
Intermediate occupations	8.79	*8.8*	*8.5*	14.29
Small employers and own account workers	10.1	*8.8*	*17*	6.53
Lower supervisory and technical occupat	5.96	*6.1*	*5.3*	8.22
Semi-routine and routine occupations	11.15	*8.7*	*24.1*	16.67
Never worked	0.63	*0.2*	*3.2*	1.67
Respondent's highest educational attainment (%) ***	**ALL foreign**	***mixed only***	***fully foreign***	**Native**
Degree or equivalent	46.03	*47.3*	*39.7*	29.11
Vocational qualification below degree	27.94	*29.1*	*22.2*	39.08
Higher Grade or equivalent	7.87	*7.3*	*10.6*	7.44
Standard Grade or equivalent	7.54	*8.6*	*2.3*	16.51
Other	2.17	*1.3*	*6.4*	0.12
No Qualifications	8.45	*6.3*	*18.9*	7.74
Equivalised income (%)***	**ALL foreign**	***mixed only***	***fully foreign***	**Native**
Bottom Quintile (<11,875)	18.78	*15.5*	*37.8*	20.39
2nd Quintile ($\leq$ 11,875 < 19,444)	19.03	*18.95*	*19.5*	20.41
3rd Quintile ($\geq$ 19,444 < 25,625)	18.56	*19.3*	*14.1*	19.04
4th Quintile ($\geq$ 25,625 < 37,500)	22.34	*23.4*	*16.1*	20.78
Top Quintile ($\geq$ 37,500)	21.29	*22.8*	*12.4*	19.38
Area of living by deprivation (%)	**ALL foreign**	***mixed only***	***fully foreign***	**Native**
20% least deprived	27	*28.20*	*21.25*	20.66
20-40% least deprived	19.96	*22.19*	*9.11*	21.68
40-60% least deprived	19.67	*21.26*	*11.95*	20.68
60-80% least deprived	17.16	*15.53*	*25.24*	16.83
20% most deprives	16.2	*12.82*	*32.45*	20.15

Note: ALL foreign families is a pooled sample of mixed and fully foreign families. This sample is used for most regression analyses. Mixed families are those in which only one parent was born outside of the UK. In fully foreign families both parents were born outside of the UK. In native families both parents were born in the UK.
Stars here indicate whether the distributions differ statistically across the groups.
Lack of stars suggests that the differences are not statistically significant. Significance levels: *** p<.01, ** p<.05, *p<.1

3.4.2 Representativeness of families with at least one foreign parent

The portrait of a *mixed family* from the previous section reflects, or at least does not contradict, what we know about immigrants to Scotland. The information about immigrants to Scotland is rather limited. The majority of studies focus on the UK in general, without singling out specific countries (e.g. Rienzo (2013)). Scotland-specific studies mostly provide information about the distribution and flows of immigrants to Scotland (Allen, 2013) or the labour market outcomes of immigrants (Vargas-Silva (2013a), Vargas-Silva (2013b)), although Eirich (2011) sheds light on characteristics of migrants to and from Scotland, drawing on various UK data sources. The most comprehensive source of information is the 2011 Census, results of which are being gradually released (National Records of Scotland (2013b), National Records of Scotland (2013a)). Even then, however, very little can be inferred about migrant families as its main focus is to report the migrant stock in various areas of Scotland, migrants' education levels and labour market outcomes. It does encompass the entire legal migrant population resident in Scotland at the time of the Census, but does not (as yet) provide detailed information on migrants' family situation. According to Eirich (2011), 23% of foreign-born residents of Scotland were living in a family with a child. Hence, only about a quarter of the migrant Census respondents constitute a potentially comparable group to the GUS respondents. It must be noted, however, that in *mixed* families usually just one of the parents was born abroad which further complicates any comparisons. Importantly, Census data capture the situation in Scotland in 2011 and the group participating in GUS must have been residing in Scotland already in 2005 when the project started and beforehand, given the fact that over 98% of children in the sample were born in Scotland[4]. Therefore, any comparisons are very rough.

Nonetheless, according to the 2011 Census, 7% of Scottish residents were born outside of the UK and 5% of children in GUS data have at least one parent born outside of the UK. Further, according to the Census, almost 6% of Scottish residents spoke a *foreign* language at home[5] - exactly the same proportion as in the data I rely on.

The migrant group in the data also seems to approximately match the Scottish migrant population in terms of their socio-economic characteristics. For example, looking at NS-SEC classification of migrants, both males and females are concentrated in the lowest paid (18.2%) and in the two highest paid occupational categories (32.5%) (Vargas-Silva, 2013b). In GUS, respondents and their partners are mostly represented in the professional category (42% and 51% respectively). Still, 22% of respondents and 17% of their partners work in semi-routine and routine occupations. One could argue that the polarisation is less visible in my data, but this may be due to the fact that respondents in GUS are likely to be a specific group of migrants - middle aged, with children, potentially further into their career. Moreover, recent migration from A8 countries following the EU enlargements (2004 onwards) changed the composition of migrant stock in Scotland. The shift may have not been captured in GUS, but is becoming visible in the Census.

Similarities are also visible in terms of education with 50% of recent migrants and 33% of migrants in general in the Census having a degree qualification, compared with 46% of foreign-born respondents in GUS. Moreover, Docquier and Marfouk (2006) estimate that in 1990 40% of the migrants living in the UK had tertiary education. The number reached 49% in year 2000. Although the result is not Scotland-specific, it is in line with what I find in the data.

Despite the limitations[6], there are some indications that the group of *mixed families* may

[4]Only 28 children interviewed in wave 1 were born outside of the UK and only 108 were born in other countries in the UK.

[5]In particular, 5.56% of Census respondents aged 3 and over spoke language other than English, Gaelic or Scottish at home. Bear in mind, however, that GUS data does not necessarily exclude Gaelic and Scottish from the "foreign language" category.

[6]1) limited studies on Scotland, 2) statistics come from various data sources, 3) no focus on migrant or mixed

be representative of the migrant population in Scotland. Their size and percentage speaking foreign language is as expected and they seem similar to migrants in Scotland overall in terms of their education. Larger discrepancies emerge in NS-SEC classification but this may be due to the age structure and professional experience of the group.

3.4.3 Parental investments in children

Given the richness of the data, it is possible to shed light on the activities children living in Scotland engage in and investigate whether differences exist in broadly defined *investments* parents make in children between *native* and *mixed* families.

Investments in this context encompass any activities parents involve in with children - educational, physical, social. The idea is to see whether a child's general environment differs in terms of their exposure to various factors which may contribute to development in early years. I see it as a key element of *nurture*, which may be correlated with parents culture, hence contributing to human capital accumulation of the child (Keane and Fiorini, 2012). These *investments* may be a result of conscious choices parents make to ensure a child's development or a reflection of their lifestyle, irrespective of their family status.

I consider unconditional differences between children from mixed and native families in every day activities they are involved in. The differences are taken over percentages of respondents from *mixed* and *native* families stating that they engage in a given activity. Overall, as can be seen in Table 4 and in various plots in Appendix C.1, families participate in similar kinds of activities and with a comparable frequency, particularly with respect to physical activity and outdoor play. *Foreign* respondents do, however, on average visit friends with children less frequently than *native* parents do. In particular, almost 6% less *foreign* respondents visit friends with children most days when the child is 2 years old, but the gap closes gradually and by the age of 6 the difference is only 2% and is statistically insignificant. Children in *mixed* families are less frequently involved in educational activities such as reading books or practising rhymes and songs. The opposite, however, is true for learning shapes and letters, which almost 6% more *foreign* than *native* respondents report doing with a child every day or most days. The difference is not statistically significant though. Further, children in *mixed* families watch less TV on average, particularly as they grow older.

Differences also emerge in types of entertainment outside home that parents provide for their children (see: Table 5 and Appendix C.1). For example, depending on child's age, 8-12% more respondents from *mixed* families have taken their children to a library and 6-16% to museum *in the previous year*, relative to *native* families. On the other hand, a higher percentage of respondents from *native* families have been to the swimming pool or zoo.

Children in both groups were equally involved in various forms of physical activity *in the week prior to the interview*. By the age of 5, a lower proportion of children from *mixed* families rode a bicycle. Yet, they were more likely to have done other active sports. Hence, these differences are not necessarily indicative of significant variations in lifestyle. Children may simply prefer other types of activities.

families, 4) many outcomes not comparable and differently defined

Table 4: Differences in parental investments in children between mixed and native families (Δ %)

Questions: How often do you or your partner.... ?															
	take child to park			**visit friends with children**					**read books to child**	**play outdoor games**		**paint or draw**	**sing, recite with child**		
Age of child	3	5**	6	2**	3**	4	5**	6	1	3***	1	3**	3**	1*	3**
Every day / most days	6.41	2.98	-0.57	-5.69	-4.45	-3.56	-2.1	-2.44	0.3	-4.97	-0.37	-1.9	1.53	-3.7	-9.55
Once or twice a week	-5.29	-5.37	0.63	-0.08	-2.16	-1.58	-2.7	-0.66	-3.17	-0.21	0.73	0.27	0.46	1.5	6.62
Once a fortnight	-1.01	-4.62	-3.91	-1.03	2.5	0.45	-1.26	-0.6	0.02	3	0.32	-0.61	0.93	0.52	0.79
Once every 1 or 2 months	-0.88	2.08	1.01	4.4	1.18	2.42	5.87	4.71	0.2	-0.33	-0.13	0.65	-0.78	0.41	0.23
Once every 3 or 4 months	-0.11	0.75	0.53	1.35	0.81	0.38	-0.74	0.41	-0.34	0.25	-0.02	-0.08	0.06	-0.13	0.29
Once every 6 months	-0.27	1.03	0.68	0.26	-0.05	0.35	2.63	0.5	0.05	-0.12	-0.02	0.36	-0.38	-0.06	0
Once a year or less often	-0.19	-0.26	0.43	-0.27	0.61	0.08	0.72	-0.35	-0.1	-0.12	-0.02	-0.27	0.25	0	0.36
Varies too much to say	0.5	0.81	0.32	0.57	1.17	0.36	-1.06	0.38	-0.27	1.37	0.03	-0.02	0.05	0.25	0.78
Never	0.84	2.59	0.88	0.5	0.39	1.09	-1.35	-1.98	3.32	1.12	-0.52	1.6	-2.12	1.21	0.49
Variable	**learn letters, shapes**	**use computer**	**take child to library**			**TV watched (weekdays)**						**TV watched (weekends)**			
Age of child	3	3	1**	3*		2	3	4	5***	6**	2	3	4	5	6
Every day / most days	5.85	-15.4	0.32	-0.35	None	0.57	0.54	-0.12	2.17	2.5	0.72	-0.43	-0.24	1.17	1.93
Once or twice a week	-5.12	11.74	1.55	3.33	Up to 30 minutes	-2.81	-1.28	0.35	0.8	-0.82	-2.99	-0.91	0.63	1.86	1.97
Once a fortnight	-0.27	1.06	2.74	2.4	30 minutes to 1 hour	0.08	2.66	0.96	4.15	3.12	1.19	1.49	0.69	0.07	-0.27
Once every 1 or 2 months	0.18	0.05	-0.01	1.6	1 to 2 hours	-0.05	-3.97	4.29	-1.79	0.82	-0.41	-0.12	5.58	-2.99	-0.65
Once every 3 or 4 months	-0.42	0.71	0.89	-0.01	2 to 3 hours	0.89	1.11	-4.6	-2.54	-3.13	1.71	-1.88	-3.06	-3.27	3.19
Once every 6 months	-0.06	0.06	-0.16	1.33	3 to 4 hours	0.71	0.75	-1.21	-1.35	-1.17	-0.02	0.92	-0.44	3.7	0.73
Once a year or less often	-0.36	0.72	0.94	-0.65	4 to 5 hours	0.64	1.08	0.63	-0.65	-1	-0.7	0.86	-2.04	1.35	-2.99
Varies too much to say	-0.53	1.31	0.3	1.7	5 or more hours	-0.05	-0.89	-0.3	-0.8	-0.33	0.5	0.06	-1.12	-1.9	-3.94
Never	0.72	-0.23	-6.58	-9.37											

Note: Here Δ is a raw difference in percent of respondents from mixed and native families reporting that they engage in a given activity.
Note: Stars next to the age indicator show whether the distributions differ statistically for mixed and native families for given age of the child. Lack of stars suggests that the differences are not statistically significant.
Significance levels: *** p<.01, ** p<.05, *p<.1

Table 5: Differences in types of entertainment experienced (Δ %)

Questions: Child has been to ..... since last year.												
	library			concert, play			swimming pool			sport event		
Age	2	4	6	2	4	6	2	4	6	2	4	6
	12.45***	6.56***	8.64**	-0.59	-6.17***	-2.23	-13.69***	-9.83***	-8.3***	-1.57	-4.43**	-0.06
	museum, gallery			zoo, aquarium			cinema			religious event		
Age	2	4	6	2	4	6	2	4	6	2	4	6
	15.82***	6.52***	10.04***	-2.77	-3.3**	-6.6*	4.27***	-3.85*	3.13	2.8	6.27***	1.67

Note: Here Δ is a raw difference in percent of respondents from mixed and native families reporting that they engage in a given activity.
Stars indicate whether the differences are statistically significant. Significance levels: *** p<.01, ** p<.05, *p<.1

Table 6: Differences in physical activity in children in native and mixed families (Δ %)

Questions: Child ..... last week.								
	rode a bicycle		**kicked a ball**		**danced**		**ran/jumped**	
Age	3	5	3	5	3	5	3	5
	-0.84	-5.69***	1.09	-1.1	2.34	0.07	0.56	-0.25
	swam		**played in play area**		**played in park**		**did other active sport**	
Age	3	5	3	5	3	5	3	5
	-2.37	0.65	-4.59*	-4.35*	-3.68	-2.69	3.49	4.07**

Note: Δ is a raw difference in % of respondents from mixed and native families reporting that they engage in a given activity.
Stars indicate whether the differences are statistically significant.
Significance levels: *** p<.01, ** p<.05, *p<.1

The observations suggest heterogeneity in families' lifestyles, which may be a reflection of parents' lifestyles in general, e.g. whether they are physically active or have passion for literature, irrespective of having a child and be correlated with their socio-economic characteristics. They may, equally, be a result of conscious decisions made by parents regarding their children's upbringing. In particular, parents in *mixed* families may spend more time with their children practising letters as they feel a need to do so, given that children in many cases are bilingual and are learning two languages simultaneously.

3.4.4 Parental views and ambitions

Parents have distinctive ambitions for their children and views regarding upbringing. A higher percentage of parents in *mixed* families hope for their child to complete a postgraduate degree. The difference may not be so surprising, bearing in mind that it is unconditional and that a higher percentage of *foreign* respondents have completed tertiary education. In line with this finding, relative to native parents, 6% less foreign parents wish that by mid-20s their child had a full-time job and 3% more wish for the child to have gone to university.

Greater differences between respondents from *mixed* and *native* families are visible in their attitudes towards parenting. Specifically, native respondents were more likely to say that they agree or strongly agree that nobody can teach them how to be a good parent, although the gap narrows with the age of the child and becomes insignificant by the age of 4. Equally, a lower percentage of native parents agree that it is important to stick to a routine with children. A difference emerges also in the view that it is better for children to have two parents than one where between 17% and 19% more *foreign* than *native* respondents agree or strongly agree with the statement. At the same time, as can be seen in Table 8 and plots in Appendix C.1,

respondents from *mixed* families are less likely to have used disciplining techniques, such as time out, naughty step or ignoring bad behaviour with the child. They are also less likely to say that they smack the child or use a raised voice.

These contrasts in opinions may be partly a reflection of the family situation, with a higher percentage of *native* households being lone parent families. They may also suggest that families differ in their approach to upbringing on difficult to measure dimensions. There is potential for this heterogeneity to translate into child's outcomes, particularly in sphere of non-cognitive skills and behaviour (Borghans, Lee Duckworth, et al. (2008), Carneiro, Crawford, et al. (2007), Heckman and Rubinstein (2001)).

Table 7: Differences in parental attitudes and ambitions for the child (Δ %)

	How far do you see child's education?		**Wish the child by mid-20s had/was....**	
Age	5	6***		6
To get Standard Grades	-0.28	-2	gone to university	3.49***
To get Higher Grades	0.41	-3	a full-time job	-6.19***
To finish one or two year college course	-4.58	2	in family business	1**
To finish three or four year degree course	6.03	-1	a full-time volunteer	4.83***
Masters, PhD, medical or other postgraduate	6.23	5	a family	2.49***
I don't really mind	5.84	-1	left home	-2.08
			been travelling	4.01**
			done something else	-0.023

Nobody can teach you how to be a good parent

Age	1***	3*	4	6
Strongly agree	-8.46	-3.91	-1.93	0.78
Agree	-4.5	-3.96	-5.52	-4.67
Neither agree nor disagree	2.14	2.2	1.88	1.25
Disagree	9.23	3.54	5.63	2.09
Strongly Disagree	1.6	2.14	-0.06	0.55

It is more important to go with what the child wants than to stick to a routine

Age	1***	3***	4**	6***
Strongly agree	0.87	-0.52	0.59	-0.89
Agree	6.51	6.18	6.59	10.04
Neither agree nor disagree	1.27	6.04	-4.07	1.23
Disagree	-8.04	-10.95	-1.02	-5.89
Strongly Disagree	-0.6	-0.75	-2.09	-4.48

It s better for children to have two parents than one

Age	1***	3***	4***	6***
Strongly agree	15.74	18.48	14.87	10.41
Agree	3.02	-0.24	3.74	7.12
Neither agree nor disagree	-6.32	-6.57	-6.38	-3.57
Disagree	-10.43	-10	-10.12	-12.09
Strongly Disagree	-2	-1.66	-2.1	-1.87

Note: Δ is a raw difference in % of respondents from mixed and native families reporting that they engage in a given activity.
Stars next to Δ indicate whether the differences are statistically significant. Stars next to age indicate whether the distributions of responses differ.
Significance levels: *** p<.01, ** p<.05, *p<.1

3.4.5 Child outcomes

I start by comparing the performance of children in the two groups, without conditioning on any other variables, to answer the question whether children from *mixed* and *native* families differ in outcomes. From Table 9 it is clear that children from *mixed* families score lower in the Vocabulary Naming exercise at both age 3 and 5 and the difference is statistically significant, but the gap between the average score of the two groups narrows with age. The same cannot

Table 8: Differences in use of disciplining techniques (Δ %)

Used with child....			
Age	2	4	6
time out	0.12	-3.5*	-8.34**
rewards	3.93***	1.27**	0.39
ignored bad behaviour	-7.39	-4.38*	-2.1
smacking	-0.43	-4.06**	-6.19***
naughty step	-6.6***	-7.17***	-4.87
raised voice	3.16	-0.23	-5.4**
removing treats	-7.69***	-6.02***	-2.48

Note: Here Δ is a raw difference in outcomes between native and mixed families.
Stars indicate whether the differences are significant.
Significance levels: *** p<.01, ** p<.05, *p<.1

be said about the Picture Similarities scores, where there are no significant differences between the two groups; if anything, *mixed* family children seem to overtake *native* children by the age of 5, but the difference is minimal.

There are also visible differences in the percentage of children who, according to the respondents, can be *mostly* understood by strangers, but this gap also closes with age and disappears entirely by the age of 5. A similar pattern emerges for the child being understood by family and friends, but not for the understanding of a child by the respondent.

Children perform comparably in non-cognitive and behavioural assessments, such as CSBS and SDQ. This observation holds for the total scores, as well as their composites (see: Table 9). However, the difference in CSBS total score (-.905 or 17% of standard deviation), is statistically significant; it is due to the difference in performance of children in CSBS social and symbolic composite part of the test. Statistically significant differences also emerge in the peer relationships component of SDQ score.

I consider measures of motor and physical development and find no differences across the two groups. The results of this analysis can be found in Table 23 in Appendix C.

To summarise, based on the unconditional comparisons, children from *mixed families* lag behind in cognitive outcomes that are most likely driven by language skills (i.e. speech-related) and not any other aspects of development. In particular, there seem to be no differences in non-cognitive and behavioural outcomes. Judging by the average Picture Similarities score, the cognitive skills are also not affected (at least to the extent measured by the test). Moreover, the differences in speech-related measures narrow down with age of the child, suggesting that any early years disadvantage is being gradually remedied.

My hypothesis that the differences in outcomes may be due to the language the child speaks finds support in the results of the speech assessment provided by respondents themselves. When asked about understanding of the child by strangers, respondents in *mixed families* are less likely to say that child is understood but it is no longer the case when it comes to being understood by the respondent himself. This is intuitive. If the result is genuinely driven by the fact that the child is not such a capable English speaker, then family should still understand it as they usually speak the same language.

Table 9: Summary of the outcomes variables

CSBS scores (age 2, birth cohort only)					**SDQ score**				
		N	mean	st.dev.			N	mean	st.dev.
CSBS total score	native	3801	49.312	5.325	SDQ total score	native	16746	7.587	4.642
	mixed	348	48.408	5.295		mixed	1542	7.785	4.642
	difference		0.904***			*difference*		.197*	
CSBS social composite					*By age*		4	5	6
	native	3861	22.571	2.667	native		7.846	7.647	7.239
	mixed	357	22.207	2.700	mixed		7.894	7.906	7.531
	difference		-.364***		*difference*		.047	.258	.291
CSBS speech composite							N	mean	st.dev.
	native	4057	11.273	2.596	SDQ emotional score	native	16833	1.231	1.502
	mixed	394	11.195	.129		mixed	1561	1.313	1.517
	difference		- .077			*difference*		.081***	
CSBS symbolic composite					*By age*		4	5	6
	native	4027	15.357	1.773	native		1.178	1.245	1.275
	mixed	397	14.881	1.947	mixed		1.232	1.369	1.346
	difference		- .475***		*difference*		.054	.124**	.070
	BAS scores						N	mean	st.dev.
		N	mean	st.dev.	SDQ conduct score	native	16838	1.784	1.463
Picture Similarities	native	7009	61.301	28.718		mixed	1562	1.722	1.462
	mixed	658	62.613	29.196		*difference*		- .062	
	difference		1.312		*By age*		4	5	6
By age		3	5		native		1.978	1.770	1.586
native		51.093	72.130		mixed		1.921	1.709	1.505
mixed		51.656	74.187		*difference*		- .056	-.060	-.080
difference		.563	2.056*				N	mean	st.dev.
		N	mean	st.dev.	SDQ hyperactivity score	native	16808	3.539	2.278
Vocabulary Naming	native	6999	65.451	30.122		mixed	1552	3.552	2.289
	mixed	654	58.750	33.300		*difference*		.012	
	difference		6.700***		*By age*		4	5	6
By age		3	5		native		3.588	3.605	3.416
native		56.990	74.386		mixed		3.492	3.619	3.548
mixed		48.420	69.598		*difference*		-.095	.014	.132
difference		-8.569***	-4.787***				N	mean	st.dev.
					SDQ peer relationships	native	16805	1.053	.010
						mixed	1559	1.226	.036
						difference		.172***	
					By age		4	5	6
					native		1.136	1.040	.975
					mixed		1.311	1.218	1.137
					difference		.174***	.178***	.161***
							N	mean	st.dev.
					SDQ pro-social score	native	16823	8.132	1.702
						mixed	1562	8.056	1.712
						difference		-.075*	
					By age		4	5	6
					native		7.849	8.176	8.399
					mixed		7.778	8.129	8.299
					difference		-.071	- .046	- .100*

Note: Differences are taken between mixed and native children. Statistical significance: *** p<.01, ** p<.05, *p<.1

Table 10: Raw differences in respondent-assessed speech outcomes (Δ %)

Child can be understood by...				
		strangers		
Age	2*	3***	4***	5
mostly	-4.16	-6.88	-3.65	0.48
sometimes	-0.06	1.42	1.28	0.27
rarely	4.22	5.46	2.37	-0.75
		family and friends		
Age	2*	3**	4*	5
mostly	-4.72	-3.61	-1.98	2.17
sometimes	2.23	2.39	2.47	-1.69
rarely	2.49	1.23	-0.49	-0.21
		the respondent		
Age	2	3*	4	5
mostly	-1.22	-1.85	-0.61	1.05
sometimes	0.58	1.8	0.74	-0.76
rarely	0.64	0.05	-0.12	-0.29

Note: Here Δ is calculated as a difference between outcomes of children in mixed vs native families. Stars next to age indicate whether the distributions of responses differ.
Significance levels: *** p<.01, ** p<.05, *p<.1

4 Conditional comparisons

The unconditional analysis suggested that, although there are some differences between families in terms of socio-economics and in parental views, parents often engage in similar activities with children and children perform comparably in majority of dimensions, except for those related to speech. I hypothesise that the difference in the latter is driven by the fact that high percentage of children in *mixed families* are bilingual or speak a language other than English.

Comparisons without taking into account any socio-economic circumstances of the families are likely to produce a misleading picture, particularly given the fact that children's outcomes are often correlated with parental education levels (Black, Devereux, et al., 2005), income and social status (Meier Jaeger, 2011) as well as environmental factors, such as the number of siblings, social interactions the child is exposed to, etc. (Heckman and Conti (2012), Hartas (2011), etc.). It is vital to control for these factors to isolate the effect related purely to the *mixed* family status. The raw comparisons do not control for heterogeneity within the groups, whereas distinctions may emerge given specific circumstances.

Conditioning on variables key for child's performance may also, at least partly, mitigate the effect of selection of migrants. Specifically, given the presumption that *mixed* families are positively selected on socio-economics, I would like to control for the selection. Further, positively selected *foreign* parents may realise the disadvantage their children are at and consciously attempt to compensate for it. An example of such compensation could be the higher frequency with which they practice letters with children or visit the library. Since such types of investments in children matter for their cognitive development (Keane and Fiorini, 2012), they need to be accounted for in the analysis.

4.1 Empirical specification

4.1.1 Baseline

I start the analysis from an OLS regression of various outcomes, controlling for socio-economics of the household, activities parents engage in at home, physical activity and parenting methods

as proxied by attitude to discipline. The regression equation becomes:

$$Y_{it} = \alpha + \beta_1 mixed_i + \beta_2 gender_i + \beta_3 age_{it} + \beta_4 X_{it} + \beta_5 W_{it} + \gamma_t + \epsilon_{it} \quad (1)$$

where Y_{it} are various outcome measures for child i at time t, *mixed* is a dummy variable equal to 1 if at least one of child's parents is born outside of the UK, *gender* is a dummy variable equal to 1 if the child is female, *age* is a variable reflecting child's age in years, X_{it} contains household characteristics such as number of siblings, whether family full, parental education level and NS-SEC classification, geographical location by the index of deprivation, and W_{it} contains the variables directly related to child's upbringing such as activities child engages in at home (rhymes, letter and reading practice, use of computer, watching TV), physical activity (play outside, running, jumping, etc), outings (visits to library, museum, zoo, gallery, swimming pool, cinema) and discipline (use of naughty step, time out etc.) Where the outcome variable was measured at more than one point in time, I cluster standard errors at an individual level and include time fixed effect γ_t.

The measures considered here are the BAS outcomes (Picture Similarities and Vocabulary Naming score), Strength and Difficulties Questionnaire and Communication and Symbolic Behaviour Scale. Given that BAS outcomes are recorded on a percentile scale and SDQ and CSBS scores, even though discrete, have a rather wide range (i.e. 0-34 and 0-57), I see OLS as an appropriate regression method. I also briefly look at the respondent-assessed speech development of the children, which is measured using an ordinal variable (1-3). For this outcome I rely on OLS and ordered probit but do not report all results, as the relationship is insignificant once controls are included.

Given my hypothesis that any effects found in the regressions are likely to be driven by the fact that many children in the *mixed* families are bilingual, I: 1) use language as an alternative indicator of *mixed* family status and rerun the regressions (see: Eq. 2) to then 2) introduce an interaction between language spoken at home and native/mixed family status as control variables. The regressions become:

$$Y_{it} = \alpha + \beta_1 language_i + \beta_2 gender_i + \beta_3 age_{it} + \beta_4 X_{it} + \beta_5 W_{it} + \gamma_t + \epsilon_{it} \quad (2)$$

$$\begin{aligned} Y_{it} = \alpha + \beta_1 mixed_i + \delta language_i + \theta mixed_i \times language_i \\ + \beta_2 gender_i + \beta_3 age_{it} + \beta_4 X_{it} + \beta_5 W_{it} + \gamma_t + \epsilon_{it} \end{aligned} \quad (3)$$

All the controls remain unchanged and language is a dummy variable equal to 1 if a child speaks English and another language at home and to 0 if child only speaks English at home. I exclude monolingual speakers of another language from the regression (n=68) as combining bilingual children with monolingual speakers of language other than English is problematic, as the children are likely to face different challenges. Bilingual children learn two languages simultaneously, but when they master them, they are fluent in English and hence their interaction with other members of the society is eased. Children who only speak another language are likely to face a new set of difficulties upon beginning school when they need to learn English. The group of monolingual speakers of another language is negligible in the data and their exclusion from the regression does not change the results.

4.1.2 Differential impacts

I analyse further the outcomes for which I find an effect of being in a mixed family or being bilingual. In particular, I am interested in gender and age-variation in performance. It is reasonable to think that girls may develop differently from boys, also in the context of bilingualism and multiculturalism. Given the observation in bilingualism literature (Baker, 1999) that, although bilingual children are at a disadvantage in certain areas of development in early years,

they catch up with or even supersede their peers and the fact that in unconditional comparisons gaps seem to narrow with age, I look at changes in the difference with age. Hence, I introduce further interaction terms of *mixed* or *language* with age and gender into the regressions.

$$Y_{it} = \alpha + \beta_1(mixed/language)_i + \beta_2 gender_i + \theta(mixed/language)_i \times gender_i + \beta_3 age_{it} + \beta_4 X_{it} + \beta_5 W_{it} + \gamma_t + \epsilon_{it} \quad (4)$$

$$Y_{it} = \alpha + \beta_1(mixed/language)_i + \beta_2 gender_i + \theta(mixed/language)_i \times age_{it} + \beta_3 age_{it} + \beta_4 X_{it} + \beta_5 W_{it} + \gamma_t + \epsilon_{it} \quad (5)$$

4.2 Results

4.2.1 Cognitive outcomes

In Table 11 I present results of the baseline regressions for cognitive outcomes, i.e. BAS scores and respondent-reported speech assessment. Looking at Panel A, it is clear that both being a child in *mixed* family and speaking language other than English at home lowers the Vocabulary Naming Score. The results do not change upon inclusion of any controls. The coefficients in columns (1) to (6) of Table 11 suggest that a child from a mixed family scores on average between 6.24 and 6.79 percent lower than a child from a native family, depending on regression specification. This is a big impact equal to 9.6-10.4% of the mean score and 20.8-22.6% of the standard deviation.

Using language as an explanatory variable produces bigger impacts; bilingual children score between 13.5 and 15% lower than English speaking monolingual children, which may be indicative of language acting as a channel for the difference. Upon inclusion of the interaction term between family status and language, the coefficients on language and mixed are no longer significant, which can be expected given the high correlation between the variables. The coefficient on the interaction term, suggests that bilingual children in mixed families score on average 13.47% lower than mixed monolingual children, highlighting the role of language.

The results for BAS Picture Similarities test (Panel A, Table 11) are statistically insignificant and negligible in size, confirming the previous observation that the cognitive skills of children may not differ across the two groups.

I also consider impacts on the ability of a child to be understood by strangers and report results in Panel B of Table 11. Despite initial statistically significant impacts, any effect of being in a mixed family disappears upon inclusion of control variables. When language is used as an explanatory variable, the effect remains marginally statistically significant, again highlighting the role of language. I relate the less conclusive findings to the fact that the measure is subjective and depends on parental perception of what being understood means. Nonetheless, a positive coefficient would suggest that a child from a *mixed* family or who speaks a different language is less likely to be understood by strangers[7]. I repeat a similar analysis for the two remaining questions in the respondent-assessment measure of speech development but find no significant results. The output can be found in Appendix D.

4.2.2 Non-cognitive and physical development

I analyse the non-cognitive development by first looking at the total scores for SDQ and CSBS (see: Table 12). The coefficient on SDQ score is positive, suggesting a slightly worse performance of children in *mixed* families, but becomes statistically insignificant upon inclusion of

[7]Note that the speech-assessment variable was coded in the following way: 1 - often, 2 - sometimes, 3 - rarely or not at all. Therefore, a positive coefficient in ordinal probit regression indicates that a child is less likely to be understood by strangers.

any controls, confirming what was clear also in summary statistics, that the two groups score comparably in SDQ.

The total CSBS score is affected by child's status. According to columns (1) to (5) of Panel C, a child from a *mixed* family scores on average between .854 to .930 less than other children, which is equivalent to 1.7-1.9% of the mean score but as much as 16-17.45% of the score's standard deviation. This is also very close to the unconditional difference discussed before. A larger impact is found for bilingual children. However, in the regression including an interaction term between *mixed* and language, I find only the coefficient on *mixed* to be statistically significant, undermining the role of language in this score.

Where could this difference in impacts between SDQ and CSBS be coming from, given that they both measure non-cognitive aspects of child's development? Unlike the SDQ test, the CSBS test was only taken at one point in time (age 2) and on one cohort of children (BC). It is possible that differences were more visible at this stage, but it is difficult to assess robustness of this result given the cross-sectional nature of the outcome. I cannot investigate whether the performance changes with age either. The Strength and Difficulties Questionnaire was used to assess children from the age of 4 onwards and, although it measures similar aspects of non-cognitive development, questions respondents were asked differ from those involved in CSBS analysis making comparisons infeasible. It is possible, however, that by the age of 4 children improve their performance and hence SDQ tests do not reveal any differences.

This initial analysis does not provide firm conclusions regarding non-cognitive development of children. The nature of the tests and the arbitrary way in which total scores are obtained (by summing up the composite scores), may raise questions about validity of the findings and whether some existing differences become invisible due to aggregation. I replace the total scores with clusters of SDQ and composite scores of CSBS as dependent variables and run separate regressions for these elements of assessments only. The results of fully specified regressions, including all previously used controls, can be seen in Panel B and Panel D of Table 12.

Among the subcomponents of the total SDQ score, the only one which is affected by the mixed family status or language is the *peer problems score*. The regressions imply that it is .117 higher for children from mixed families and .306 higher for bilingual children, but there does not seem to be any significant interaction between language and mixed family status. This is a non-negligible effect equal to 8.5% of standard deviation of the score due to the mixed family status and 22.2% of standard deviation due to bilingualism. As is outlined in Appendix A.5, the peer problems score is calculated on the basis of questions regarding the child having friends, liking other children, being bullied by other children and getting on better with adults than children. The effect on the peer problems score is channelled through the child being picked on and getting on better with adults (see regressions in Appendix D).

Other single elements of the total SDQ score which children from mixed families and bilingual children seem to perform worse in include being considerate of other people's feelings, sharing with other children and fighting with other children (see detailed analysis by question in Appendix C). It is possible that these elements of relationships with peers are influenced by language and ability to communicate.

It is clear from Panel D of Table 12 that the factor driving the result on the total CSBS score is the symbolic composite. The component is aimed to capture children's understanding of words (reaction to own name, understanding of phrases) and object use (appropriate use of objects, ability to stack blocks, interest in playing with objects and pretend playing with toys). The results suggest that a child from a mixed family scores .491 lower in the symbolic component which is equivalent with 3.2% of the mean score and 27% of its standard deviation. The effect is slightly bigger when language is used as the main explanatory variable but, with an interaction term, it is the mixed family status which can explain the result. Both elements of the symbolic composite are negatively affected by the child's status; children from *mixed*

families know a lower number of words and are less likely to use objects appropriately and to pretend play with toys (see detailed analysis by question in Appendix C).

The data set also provides an alternative measure for CSBS, which takes a form of a dummy variable equal to 1 if child falls into the "concern group" given the assessments threshold points[8]. Using this variable as a benchmark indicator of performance may be more suitable as, although still arbitrary, it highlights a more important aspect of the assessment - whether the children are performing well below the average. The results of the regressions can also be found in Panel D and suggest that children from mixed families are more likely to fall into the concern group with respect to the social composite score but no other elements of CSBS assessment.

[8]According to Wetherby and Prizant (2001) criterion levels for concern are set at more than 1.25 standard deviation below the mean.

Table 11: Baseline regression outcomes for cognitive skills measures

Panel A: British Ability Scales Assessment

BAS Vocabulary Naming Score

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) OLS	(12) OLS	(13) OLS
mixed	-6.244***	-6.792***	-6.451***	-6.622***	-6.645***	-6.660***							-.211
	(1.565)	(1.516)	(1.505)	(1.476)	(1.474)	(1.472)							(1.766)
language							-14.979***	-13.831***	-13.387***	-13.299***	-13.438***	-13.470***	-3.422
							(2.061)	(1.999)	(1.978)	(1.941)	(1.952)	(1.948)	(3.442)
mixed*language													-13.474***
													(4.406)
Controls													
household controls	no	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	yes
activities at home	no	no	yes	yes	yes	yes	no	no	yes	yes	yes	yes	yes
outings	no	no	no	yes	yes	yes	no	no	no	yes	yes	yes	yes
physical activity	no	no	no	no	yes	yes	no	no	no	no	yes	yes	yes
discipline	no	no	no	no	no	yes	no	no	no	no	no	yes	yes
N	4190	4054	4018	3975	3975	3975	4189	4054	4018	3975	3974	3974	3974
R-squared	0.105	0.198	0.205	0.216	0.216	0.217	0.113	0.196	0.209	0.219	0.221	0.222	.224

BAS Picture Similarities Score

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) OLS	(12) OLS	(13) OLS
mixed	1.583	.211	-.070	-.266	-.332	-.284							0.216
	(1.240)	(1.253)	(1.241)	(1.232)	(1.230)	(1.229)							(1.454)
language							-0.21	-1.361	-1.329	-1.394	-1.703	-1.710	-1.333
							(1.671)	(1.743)	(1.730)	(1.725)	(1.732)	(1.726)	(2.923)
mixed*language													-.709
													(3.811)
Controls													
household controls	no	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	yes
activities at home	no	no	yes	yes	yes	yes	no	no	yes	yes	yes	yes	yes
outings	no	no	no	yes	yes	yes	no	no	no	yes	yes	yes	yes
physical activity	no	no	no	no	yes	yes	no	no	no	no	yes	yes	yes
discipline	no	no	no	no	no	yes	no	no	no	no	no	yes	yes
N	4192	4054	4018	3975	3974	3975	4191	4058	4022	3978	3977	3977	3977
R-squared	0.006	0.187	0.191	0.195	0.198	0.195	0.141	0.19	0.193	0.197	0.2	0.2	0.2

Panel B: Self-reported speech assessment - Can child be understood by strangers?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	OLS	**ordered probit**	**ordered probit**	**ordered probit**	**ordered probit**	**ordered probit**	**ordered probit**	**ordered probit**	**ordered probit**	**ordered probit**	**ordered probit**	**ordered probit**	**ordered probit**	**ordered probit**
mixed	.073***	.155***	.138***	.139***	.094	.091	.095							.042
	(.019)	(.048)	(.050)	(.068)	(.068)	(.068)	(.068)							(.091)
language								.257***	.214***	.202***	.168*	.151*	.153*	.113
								(.054)	(.064)	(.064)	(.088)	(.089)	(.089)	(0.197)
mixed*language														.013
														(.235)
Controls														
household controls	no	no	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	yes
activities at home	no	no	no	yes	yes	yes	yes	no	no	yes	yes	yes	yes	yes
outings	no	no	no	no	yes	yes	yes	no	no	no	yes	yes	yes	yes
physical activity	no	no	no	no	no	yes	yes	no	no	no	no	yes	yes	yes
discipline	no	no	no	no	no	no	yes	no	no	no	no	no	yes	yes
N	7451	7451	6791	6617	4061	4098	4058	7447	6790	6616	4061	4058	4058	4058
R-squared	0.146	0.146	0.193	0.205	0.115	0.119	0.123	0.157	0.193	0.203	0.115	0.123	0.123	0.123

Note: All regressions include gender, sweep and the cohort dummy as controls. Household controls include: number of siblings, parental education, NS-SEC household classification, equivalised income, geographical area and whether one is a single parent.Activities at home are frequency of painting/drawing, practicing rhymes, practicing letters, use of computer and TV. Outings include visits to many attractions, including museums, library, etc., physical activity includes swimming, playing in the park, running, etc. And disciplining techniques include time out and naughty step. The independent variable mixed is a dummy equal to 1 if at least one of child's parents was born outside of the UK and zero otherwise. Language is a dummy variable equal to 1 if child speaks English and another language or just another language at home. Error clustered at individual level in parentheses.
Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table 12: Baseline regression outcomes for non-cognitive assessments

Panel A: SDQ total score

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) OLS	(12) OLS	(13) OLS
mixed	.166	.359*	.361*	.217	.396*	.205							.298
	(.184)	(.188)	(.188)	(.199)	(.222)	(.183)							(.257)
language							0.675***	.474*	.570**	.474	.440	.487	0.006
							(.252)	(.253)	(.273)	(.319)	(.322)	(.319)	(0.487)
mixed*language													.381
													(.669)
Controls													
household controls	no	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	yes
activities at home	no	no	yes	yes	yes	yes	no	no	yes	yes	yes	yes	yes
outings	no	no	no	yes	yes	yes	no	no	no	yes	yes	yes	yes
physical activity	no	no	no	no	yes	yes	no	no	no	no	yes	yes	yes
discipline	no	no	no	no	no	yes	no	no	no	no	no	yes	yes
N	6683	6485	6347	5835	5658	6156	6680	6522	6346	5475	5453	5451	5451
R-squared	0.018	0.1	0.116	0.128	0.115	0.15	0.019	0.104	0.116	0.135	0.14	0.144	0.144

Panel B: SDQ subcomponents - clusters

	Emotional symptoms			Conduct problems			Hyperactivity			Peer problems			Pro-social		
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) OLS	(12) OLS	(13) OLS	(14) OLS	(15) OLS
mixed		0.083	0.088		.028	.061		-0.001	.130	.117**		.017	-.121		-.051
		(.061)	(.090)		(.061)	(.088)		(.094)	(.136)	(.059)		(.070)	(.077)		(.105)
language	.140		.110	.029		.001	.007		-.274		.306***	.168		-.034	-.097
	(.101)		(.196)	(.105)		(.162)	(.157)		(.266)		(.105)	(.188)		(.118)	(.217)
mixed*language			-.039			-.018			.263			.171			.133
			(.243)			(.223)			(.348)			(.234)			(.276)
N	5472	6156	5472	5472	6195	5472	5461	6195	5461	6156	5466	5466	6195	5471	5471
R-squared	.040	.051	.040	.098	0.114	.098	.111	.092	.111	.063	.071	.071	.064	.065	.065

Note: In this setup the regression controls include all: household controls, activities at home, outings, physical activity and discipline.

Panel C: CSBS total score

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) OLS	(10) OLS	(11) OLS	(12) OLS
mixed	-.854***	-.914***	-.900***	-.887***	-.930***							-.692*
	(.296)	(.301)	(.327)	(.326)	(.325)							(.402)
language							-1.778***	-1.728***	-1.271***	-1.058**	-1.091**	-.552
							(.465)	(.495)	(.467)	(.461)	(.461)	(.879)
mixed*language												-.076
												(1.089)
Controls												
household controls	no	yes	yes	yes	yes		no	yes	yes	yes	yes	yes
activities at home	no	no	yes	yes	yes		no	no	yes	yes	yes	yes
outings	no	no	no	yes	yes		no	no	no	yes	yes	yes
discipline	no	no	no	no	no		no	no	no	no	yes	yes
N	4149	3883	3106	3106	3106		4148	3883	3106	3106	3106	3106
R-squared	.019	.042	.123	.131	.136		.021	.044	.123	.131	.134	.134

Panel D: CSBS subcomponents - composites

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS	Whether child falls in the concern group: probit	probit	probit	probit
	Social			Speech			Symbolic			Social	Speech	Symbolic	Total score
mixed	-.207		-.208	-.244		-.197	-0.491***		-.313**	.108	.142	.379***	.153
	(.171)		(.212)	(.149)		(.188)	(.112)		(.133)	(.124)	(.091)	(.121)	(.124)
language		-.148	-0.026		-.325	-0.387		-.592***	-.124				
		(.244)	(.496)		(.201)	(.386)		(.158)	(.227)				
mixed*language			.032			.261			-.324				
			(.598)			(.482)			(.322)				
N	3144	3144	3144	3304	3304	3304	3296	3296	3296				
R-squared	0.027	.027	.027	.133	.132	.132	.170	.168	.170				

Note: In this setup the regression controls include all: household controls, activities at home, outings, physical activity and discipline.

Note: All regressions include gender, sweep and the cohort dummy as controls. Household controls include: number of siblings, parental education, NS-SEC household classification, equivalised income, geographical area and whether one is a single parent.Activities at home are frequency of painting/drawing, practising rhymes, practising letters, use of computer and TV. Outings include visits to many attractions, including museums, library, etc., physical activity includes swimming, playing in the park, running, etc. And disciplining techniques include time out and naughty step. The independent variable mixed is a dummy equal to 1 if at least one of child's parents was born outside of the UK and zero otherwise.
Language is a dummy variable equal to 1 if child speaks English and another language or just another language at home. Errors clustered at individual level in parentheses.
Significance levels: *** p<0.01, ** p<0.05, * p<0.1

4.2.3 Children with one or two foreign parents perform differently

Foreigners in mixed marriages are positively selected and differ from those in families where both parents are foreign-born. To highlight the importance of making the distinction, I repeat the introductory conditional analysis now using two distinct dummy variables, indicating that a child comes from a *mixed* or *fully foreign* family, comparing their outcomes to those of *native* children. Results can be found in Table 13. Despite the small number of *fully foreign* children, I obtain statistically significant results for some aspects of development.

As before, both groups perform worse relative to native children in Vocabulary Naming Exercise. However, children whose parents were both born outside of the UK perform significantly worse relative to natives than children with just one foreign parent. They score almost 21% lower, whilst children from mixed families score on average 4.6% lower than *native* children. The SDQ peer relationship score is also higher for children from foreign families, indicating greater problems, which is not the case for children from *mixed* families. *Fully foreign* children also perform worse in the CSBS symbolic composite score.

The results suggest that, even though only speech-related performance is affected, the negative impact is significantly greater for fully foreign children than for children from mixed families. Such an outcome may be driven by the fact that families in which both parents are foreign-born, not only are likely to use a language different from English in their every day life, but may also find it more challenging to fully assimilate in the society. They may also have worse knowledge of the institutions, educational system and hence miss opportunities to help their children successfully accommodate two (or more) cultures. I observed earlier that they are likely to be at an economic disadvantage, which means that they have less resources available to them to invest in children's upbringing. Note, however, that their children are only affected in areas where language plays a key role. They do not seem to lag behind in cognitive development in general.

Table 13: Baseline regression results of cognitive and non-cognitive outcomes for children from foreign and mixed families

	BAS Assessment		*Self-reported speech*	SDQ total score	SDQ peer relationship	CSBS total score	CSBS Symbolic
	Vocabulary Naming	Picture Similarities					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	OLS	ordered probit	OLS	OLS	OLS	OLS
both foreign	-20.78***	1.670	0.371**	0.712	0.694**	-1.138	-0.803**
	(3.963)	(3.089)	(.172)	(.622)	(.220)	(.748)	(0.318)
foreign and native	-4.560**	-0.905	0.0423	0.405*	0.0649	-0.887**	-0.440***
	(1.533)	(1.313)	(.072)	(.237)	(.068)	(.355)	(0.119)
Controls	yes	yes	yes	yes	yes	yes	yes
N	3975	*3978*	*4058*	5470	5470	3106	3296
R-squared	.221	*.199*	*.108*	.143	.071	.135	.169

All regressions include gender, sweep and the cohort dummy as controls. Household controls include: number of siblings, parental education, NS-SEC household classification, equivalised income, geographical area and whether one is a single parent.Activities at home are frequency of painting/drawing, practising rhymes, practising letters, use of computer and TV. Outings include visits to many attractions, including museums, library, etc., physical activity includes swimming, playing in the park, running, etc. And disciplining techniques include time out and naughty step. The independent variable mixed is a dummy equal to 1 if at least one of child's parents was born outside of the UK and zero otherwise. Language is a dummy variable equal to 1 if child speaks English and another language or just another language at home. Errors clustered at individual level in parentheses.
Significance levels: *** p<0.01, ** p<0.05, *, p<0.1

4.2.4 Variation by age and gender

Given that I find impacts on BAS scores, I allow them to differ by age and gender to see whether there are differential impacts in these two dimensions. The results are presented in Table 14. I find differential impacts on the outcomes of the Vocabulary Naming exercise. In particular, when including an interaction term of mixed or language with gender, it is clear that children from mixed families score over 4% lower than other children and *bilingual* children score 10%

less. Girls outperform boys, scoring 6% better on average but the gap is smaller between children from *mixed* families (girls score only 1% more) and reversed for *bilingual* children.

Children's performance improves with age and more so for children from mixed families, but not sufficiently for them to fully catch up with peers from native families by the age of 5. Children from *mixed* families score 16% lower and bilingual children almost 21% lower than native peers and they catch up at a rate of between 1.8 and 3.4% per each year of age. Note also that for the regression using language as the main explanatory variable the coefficient on the interaction term is not significant.

I find no differential impacts on Picture Similarities score which further confirms the initial finding that children score comparably in the exercise, irrespective of their background. As expected, however, the score improves with the child's age and girls score higher than boys.

When analysing the respondent-assessed child's ability to be understood by strangers, I conclude that gender does not matter for performance of children from mixed families, although girls in general are reported to be easier understood. There is a change in impact due to age; children from mixed or bilingual families are less likely to be understood by strangers but the situation improves with age, more than for native children. In fact, by the age of 5 the initial disadvantage of coming from a mixed family disappears.

I considered differential impacts by gender for SDQ and CSBS score and by age for SDQ scores[9], but found no significant effects.

[9]Note that I cannot investigate whether CSBS score differ by age as I only have a cross-sectional measure at hand.

Table 14: Outcome dynamics - differences by age and gender

	Vocabulary Naming				Picture Similarities				Can be understood by strangers			
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) OLS	(7) OLS	(8) OLS	(9) oprobit	(10) oprobit	(11) oprobit	(12) oprobit
mixed	-4.412**	-16.091***			-1.563	-3.471			.109	.656***		
	(1.878)	(4.536)			(1.600)	(4.485)			(.089)	(.217)		
language			-9.771***	-20.569***			-.554	1.785			.159*	.171*
			(2.495)	(6.128)			(2.428)	(5.659)			(.095)	(.103)
age		7.397***		6.869***		9.697***		9.499***		-.321***		-.334***
		(.365)		(.434)		(.376)		(.426)		(.021)		(.021)
mixed*age/language*age		2.374**		1.786		.803		-.879		-.149***		.0005
		(.996)		(1.418)		(1.036)		(1.340)		(.061)		(.021)
female	6.133**		6.067***		3.199***		3.317***		-.281***		-.286***	
	(.816)		(.821)		(.752)		(.745)		(.042)		(.042)	
female*mixed	-4.946*		-7.871**		2.894		-2.441		.005		.044	
	(2.962)		(3.813)		(2.470)		(3.394)		(.128)		(.105)	
N	3975	3975	3974	3974	3975	3975	3977	3977	4118	4118	4058	4058
R-squared	.216	.216	.223	.223	.195	.195	.200	.200	.127	.127	.123	.125

Note: All regressions include gender, sweep and the cohort dummy as controls. Household controls include: number of siblings, parental education, NS-SEC household classification, equivalised income, geographical area and whether one is a single parent.Activities at home are frequency of painting/drawing, practising rhymes, practising letters, use of computer and TV. Outings include visits to many attractions, including museums, library, etc., physical activity includes swimming, playing in the park, running, etc. And disciplining techniques include time out and naughty step. The independent variable mixed is a dummy equal to 1 if at least one of child's parents was born outside of the UK and zero otherwise.
Language is a dummy variable equal to 1 if child speaks English and another language or just another language at home.

4.2.5 Foreign parents can help their children catch up

Children from *mixed* families may be catching up with *native* children as they grow, but the rate is not high enough for the gap to close by the age of 5. What drives this convergence? Is it a natural process related to simultaneous acquisition of two languages or are there different returns to parental inputs into children's upbringing, depending on the family status? Do foreign parents engage in different types of activities with their children because they have group-specific returns to cognitive and non-cognitive skills?

I find that a child's involvement in various activities can explain children's outcomes to an extent, confirming the role nurture plays in a child's development (Cunha, Heckman, et al., 2010). Is this the channel through which children catch up? To shed light on this matter, I investigate further interactions of socio-economic factors, as well as indicators of various investments in children, with child's status in relation to the BAS Vocabulary Naming score. I do not present such analyses for the Picture Similarities, respondent-assessed speech development and SDQ, as found no robust impacts on these outcomes so far. I also do not analyse the CSBS score in much detail as it is a cross-sectional measure taken at the age of 2, so it is difficult to claim that any investments would have already paid off. I present the results in Table 15.

Firstly, most variables I consider matter for children's performance in the exercise. In particular, higher socio-economic classification of the family (NS-SEC categorisation, education and income) improves the score. Similar observations can be made about various activities children engage in and some, but not all, disciplining techniques. However, not all these factors have a differential impact on performance of *native* and *mixed* family children. Mixed family and bilingual children perform much worse relative to native or monolingual children on average if their parents have no qualifications. There is some weaker evidence that children from mixed families with higher incomes or who practice letters more frequently gain more than the equivalent native group. Children from mixed backgrounds seem to also benefit more from outings with parents and use of disciplining techniques. This is a reasonable finding given the hypothesis that the differences are largely driven by language, as visits to places like swimming pool or zoo present opportunities for interactions with other children and to speak English.

Table 15: Contribution of other factors to performance of mixed family children in Vocabulary Naming Exercise

Dependent variable: BAS Vocabulary Naming Exercise								
interacted variable:	NS-SEC category	degree education	no qualifications	equivalised income	frequency of drawing	singing with child	practising letters	visits to the library
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
mixed	-5.425**	-4.826**	-6.545***	-10.987***	-4.655	-9.558***	-10.887***	-8.913***
	(2.387)	(2.023)	(1.472)	(3.421)	(2.855)	(3.637)	(2.678)	(2.452)
interacted var	-1.144***	2.484**	-.758	1.574***	.658***	.755***	.849***	2.762***
	(.334)	(1.086)	(7.554)	(.366)	(.165)	(.182)	(.157)	(.772)
mixed*interacted variable	-.672	-4.124	-30.605***	1.403	-.451	.539	.943*	3.261
	(1.034)	(2.923)	(7.970)	(1.010)	(.559)	(.607)	(.501)	(2.806)
All controls included	yes	yes	yes	yes	yes	yes	yes	yes
N	3975	3975	3975	3975	3975	3975	3975	3975
R-squared	.216	.217	.217	.217	.216	.216	.217	.217
language	-11.137***	-13.256***	-13.159***	-22.535***	-11.980***	-11.994**	-18.436***	-14.514***
	(3.269)	(2.426)	(1.939)	(3.540)	(3.784)	(4.748)	(3.296)	(2.910)
interacted variable	-1.120***	2.372**	-.776	1.499***	.639***	.812***	.860***	3.027***
	(.332)	(1.074)	(7.522)	(.362)	(.163)	(.180)	(.154)	(.762)
language*interacted variable	-1.021	-.805	-24.390***	3.318**	-.303	-.250	1.187*	1.711
	(1.091)	(3.794)	(8.092)	(1.281)	(.711)	(.786)	(.620)	(3.428)
All controls included	yes	yes	yes	yes	yes	yes	yes	yes
N	3975	3975	3975	3975	3975	3975	3975	3975
R-squared	.221	.221	.221	.222	.221	.221	.221	.221

interacted variable:	concert/play	swimming pool	museum	zoo	time out	naughty step	removing treats
	OLS	OLS	OLS	OLS	OLS	OLS	OLS
	(9)	(10)	(11)	(12)	(13)	(14)	(15)
mixed	-8.222***	-15.787***	-9.342***	-13.903***	-9.530***	-8.382***	-5.373*
	(1.952)	(3.106)	(2.150)	(2.710)	(2.032)	(1.900)	(2.762)
interacted var	2.593***	2.949**	2.663***	1.328	-1.036	2.331***	.407
	(.776)	(1.200)	(.763)	(.906)	(.738)	(.821)	(.830)
mixed*interacted variable	3.441	11.247***	4.905*	9.315***	5.762**	4.667*	-1.747
	(2.469)	(3.435)	(2.734)	(3.122)	(2.429)	(2.477)	(3.070)
All controls included	yes	yes	yes	yes	yes	yes	yes
N	3975	3975	3975	3975	3975	3975	3975
R-squared	.217	.218	.217	.218	.217	.217	.216
language	-15.165***	-19.291***	-14.793***	-20.457***	-15.744***	-14.763***	-6.379*
	(2.207)	(3.122)	(2.508)	(2.842)	(2.428)	(2.348)	(3.363)
interacted variable	2.682***	2.961**	2.909***	1.361	-.897	2.516***	.683
	(.762)	(1.186)	(.752)	(.891)	(.728)	(.810)	(.820)
language*interacted variable	4.461	8.569**	2.952	10.287***	5.373*	4.253	-9.653**
	(3.511)	(3.813)	(3.641)	(3.632)	(3.070)	(3.525)	(3.879)
All controls included	yes	yes	yes	yes	yes	yes	yes
N	3975	3975	3975	3975	3975	3975	3975
R-squared	.221	.221	.221	.222	.221	.221	.222

Note: All regressions include gender, sweep and the cohort dummy as controls. Household controls include: number of siblings, parental education, NS-SEC household classification, equivalised income, geographical area and whether one is a single parent.Activities at home are frequency of painting/drawing, practicing rhymes, practicing letters, use of computer and TV. Outings include visits to many attractions, including museums, library, etc., physical activity includes swimming, playing in the park, running, etc. And disciplining techniques include time out and naughty step. The independent variable mixed is a dummy equal to 1 if at least one of child's parents was born outside of the UK and zero otherwise. Language is a dummy variable equal to 1 if child speaks English and another language or just another language at home. Errors clustered at individual level are in parentheses.

5 Limitations and discussion

I present evidence for existence of early life performance gaps between children in mixed/bilingual and native families, using data for Scotland. I show that the two types of families differ at the outset in their socio-economic characteristics, views and attitudes as well as lifestyle, measured by types of activities they engage in. Surprisingly, though, children perform comparably on an array of measures, including cognitive (Picture Similarities), non-cognitive (Strength and Difficulties Questionnaire) and motor development. Where the differences do emerge (Vocabulary Naming, CSBS and speech assessment), the outcomes are likely to be related to speech and linguistic skills. The results are similar irrespective of the definition used (mixed family vs. language) and robust to inclusion of controls. Analysing further the affected outcomes, I find a degree of convergence with age between the two groups. Its rate is insufficient for the gap to close fully by the age of 5, the last point for which data are available. Given the initial differences between the families, especially in socio-economic situation, the question arises whether the various investments parents make in children affect them differently and hence narrow or widen the gap. I find that, for mixed family children, parents' higher socio-economic status, participation in educational activities and visits to various places contribute more to their Vocabulary Naming Score than for native children. I also highlight the importance of differentiating between *mixed* and *fully foreign* families by showing that, even though children in both types of families are affected in the same way, those who grow up in households with two foreign-born parents are at a greater disadvantage at an early age. This may be because *fully foreign* families are less assimilated with the society in the receiving country and may be less able to provide the child with all necessary support to learn English.

The effects may not be causal and the implications are more likely qualitative. I rely on OLS as the main regression method and caution is needed when drawing conclusions as various forms of selection and unobserved heterogeneity may invalidate the results. The OLS regression coefficients will be biased if there is an unobserved heterogeneity among children, which is crucial for their performance and correlated with the explanatory variables already included in the regression. Given what we know about the role of various factors in explaining children's early life performance, inclusion of a rich set of controls capturing the socio-economic situation of the families, investments made in children, as well as views and attitudes of their parents, may significantly limit the extent of the problem. Naturally, factors such as ability or IQ remain excluded but they are proxied with the variables contained in the regression. The fact that the results are robust to the inclusion of further controls is also reassuring.

Various forms of sample selection pose another problem. The authors of the data set provide an extensive evidence that the participants were randomly selected from the universe of families with children of relevant age living in Scotland (Bradshaw, Tipping, et al., 2009). They also propose weights to correct for non-response and attrition in the overall sample, which I apply to all elements of the analysis. However, immigrants are a selected group from the populations in their countries of origin. This should not be of particular concern here as I compare the immigrants to natives of the receiving, rather than sending country.

Further, among those who emigrate, the more educated have a higher propensity to intermarry (Meng and Gregory (2005), Lichter and Qian (2001)), forming the *mixed* families. The positive selection may lead to bias as children of more educated parents are likely to perform better, which would close the gap between the *mixed* and *native* children. This becomes partially visible when I split the group into children with one and two foreign-born parents and a gap emerges between the two groups. However, the effect on children in *fully foreign* families is most likely different not only due to differences in socio-economics between the two groups (selection), but also due to the role cultural knowledge and assimilation play in child's upbringing. Children who have one foreign parent still are usually exposed to two cultures and languages,

which is why I find a consistent effect across the two groups for measures of development which are related to linguistic ability, but they are in a better position to adapt. Hence, I correctly identify the aspects of child's development affected by family status or bilingualism, but the size of the impact for *mixed* families may be dampened by positive selection.

If main attributes related to these two kinds of selection can be captured by the observable characteristics I have information about, the selection is explicitly controlled for in the regressions. Unfortunately, I cannot control for unobserved factors related to the choices to emigrate and to marry a foreigner, such as motivation or drive. Earlier I argued that selection into marriage is possibly not very large, as the profile of respondents in *mixed* families is similar to thar of a migrant to Scotland in general.

A higher attrition rate among the *mixed* families compared to natives is also a concern, if the characteristics related to attrition differ among the natives and foreigners. In such a case the weights proposed in the data will not correct sufficiently for the dropout rate among migrants. I discuss the problem in Appendix A.4 and argue that, although for both groups attrition is related to lower socio-economic characteristics, the differences are smaller among the *mixed* families. If the weights applied do not correct for the fact that *mixed* families with lower socio-economic status drop out of the study more, then, as with positive selection, the gap between *mixed* and *native* families will be smaller than it in fact is. Attrition among *mixed families* is closer to random, though, suggesting that the bias should be small. Moreover, the results in the paper do not change in size or scale when I do not weight the data.

I also have no information about the initial non-response rate of the group when first contacted by the project organisers. Given that weights were created on the basis of modelling which, for the first wave, only took into account respondent's age, gender and number of children in the household, the situation will be remedied only if *native* and *mixed* families responded similarly to the project.

One may ask whether there exists a suitable comparison group for mixed families and, if so, are native families best to compare to. The choice is debatable, but I argue that this is potentially the best existing group. All children participating in the study were born in Scotland and share similar environmental factors (neighbourhood, schooling, policies) from birth onwards. The main difference between them is the origin of their parents. In fact, in the majority of cases they have one British-born parent and the difference really stems from the cultural and national background of the other parent. Hence, I find them more suited for comparisons than, for example, children from foreign parents' sending countries who were subject to other institutional and cultural factors, parental nationality aside. Duncan and Trejo (2011) compare second generation immigrants from *mixed* and *fully foreign* families to assess the role of assimilation in development of skills; the extent to which I am able to follow this approach is limited as there are barely any children with two foreign parents in the GUS data.

Some of the development measures used in the analysis may be seen as subjective and hence not representative. However, the main indicators I consider, BAS assessments, are an objective evaluation of child's cognitive performance and are used in the literature as standard. Further, when used in regressions, results based on them are robust to addition of controls.

Despite these concerns, this paper provides a valuable contribution to understanding whether children from *mixed* families are at a disadvantage at an early age. It fills in the gap in our knowledge about the human capital development process by providing an insight into the evolution until the age of 5. Small children from *mixed* families perform worse, which would confirm that a performance gap does emerge early (Heckman and Conti, 2012), but only in some aspects of cognition. Moreover, parents seem to address the initial impediment and the children's skills gradually converge. They are also less disadvantaged than children with two foreign parents.

The affected outcomes are linked to linguistic ability, reinstating the role of language in the gap (Dustmann, Frattini, et al., 2010). Unlike in previous studies, however, children from

mixed families in the sample are either English speakers or bilingual, which is one of the reasons why they may catch up with peers with time. The narrowing of the gap with age is consistent with the linguistic literature arguing that bilingual children are at a disadvantage, if at all, only early on in life (Baker, 1999).

Like me, Reardon and Galindo (2009) also argue that second generation minority immigrants in the US catch up with the native children as they grow up. The finding of a closing gap is in line with the little that we know about second generation immigrants in the UK. Dustmann, Frattini, et al. (2010), looking at older children, find that the performance gap closes with age and varies across minorities.

My analysis also reinstates the importance of household income, education and investments parents make in their children (Keane and Fiorini (2012), Ermisch (2008), Hartas (2011)) and that they play a greater role for children from *mixed* families.

In their research, Duncan and Trejo (2011) suggest that, in the US, second generation immigrants from mixed marriages perform better relative to those from fully foreign families. I reach similar conclusions for children in Scotland and propose a further argument that children from *mixed* families do not lose out relative to native children. However, I focus on a broadly defined group which may be masking heterogeneity related to one's origins.

Even if qualitative in nature, this analysis constitutes a starting point on the way to defining when exactly the educational gap may be emerging, what drives the differences and what factors play role in narrowing it. It is hoped to shed light on early years' gaps and whether exposure to two cultures and languages fosters or hinders child's development. Despite the specific focus on *mixed* families, it is still informative given the growing numbers of intermarriage in Europe.

Scotland has experienced a new wave of migration since the data was collected and it would be ideal to undertake a similar analysis for children who participated in Birth Cohort 2, as their parents are more likely to be new immigrants. Their length of stay in the UK may be key for child's development since assimilation takes time and parents may lack Scotland-specific 'cultural knowledge'. The composition of the migrant group may have changed as well, mostly in terms of socio-economic characteristics.

Controlling for parental country of origin would be another extension adding an insight into the types of culture which matter for children's upbringing. Reardon and Galindo (2009) have argued that significant variation exists within Latino groups which is key for the children's outcomes. Similar considerations should apply here. However, at the moment data does not allow for such distinctions.

A Further elaboration on data and variables

A.1 Selection for the project

In preparation for the first wave a named sample of approximately 10 700 children was selected from Child Benefit records to give an achieved sample of 8 000 overall. The sampling frame was based on the geographical Data Zones for Scotland used by the Scottish Executive for purposes of releasing the small area statistics. The areas are nested within the Local Authority areas in Scotland and contain between 500 and 1000 household residents each. The zones were aggregated, sorted by Local Authority and the Scottish Index of Multiple Deprivation Score. Of those, 130 areas were selected at random and data for all children fitting the birth date criteria and living within the areas was released by the Department of Work and Pensions (DWP). Within each sample point, all eligible babies and three-fifths of toddlers were selected. Exclusions were made for 'sensitive' cases and children that had been sampled for research by the DWP in the previous 3 years. If more than one child was eligible within a household, one was selected at random.

One concern with such a selection procedure is that potentially not all families residing in Scotland register for Child Benefit, which every child under the age of 16 (or under the age of 20 if in education) is entitled to. In such a case, the initial population which was subjected to selection for the purposes of the project will not be equivalent with the universe of children residing in Scotland. This should not pose problems if we believe that those not claiming the Child Benefit do not differ from the rest of the population. It is unlikely to be the case however. It is reasonable to think that people not claiming the entitlement are either sufficiently well-off not to see a need of doing so or they are under-informed and do not know they can claim the benefit. One could argue that in the case of foreign-born citizens, many may not have sufficient knowledge of the British welfare system and be under-represented in the data. This would be contradicting the observation that the proportion of foreign born respondents in the data is similar to the overall proportion of migrants in Scotland, as registered by 2011 Census. Moreover, according to HM Revenue and Customs (2012), the uptake of the Child Benefit is persistently high, oscillating between 97% in 2006 and 96% in 2010, suggesting that the scale of the problem may be negligible. Analyses for earlier years are not available but are likely to be in line with the information cited here. In this situation, I see this limitation of the sampling procedure as a minor issue, especially given the fact that data was additionally weighted to closely match the population.

A.2 Response rates

The response rates within the first sweep reached around 80% of all in-scope. By the final sweep the response rate among those who initially participated in the study falls to 77% for natives in the child cohort and 71% for natives in the birth cohort. The attrition rate for group, I defined as *mixed*, is larger - 72% of those in the child cohort and 62% of those in the birth cohort initial sample have still participated at a final sweep.

Table 16: Sample response rates (%)

	Child cohort		Birth cohort	
	native	mixed	native	mixed
sweep 1	100 (N=2609)	100 (N=250)	100 (N=4715)	100 (N=502)
sweep 2	87.543	86.400	86.957	82.072
sweep 3	81.794	79.200	81.103	73.506
sweep 4	77.463	71.600	77.243	70.120
sweep 5			74.337	65.339
sweep 6			70.923	62.351

A.3 Weighting procedures

The data was weighted to correct for selection and attrition. Weights were created for the sample after each sweep and different weights are suggested for cross-sectional and longitudinal analyses. Il discuss only the panel weights as they are relevant for this study. For more details, please consult Bradshaw, Tipping, et al. (2009).

At every sweep except 1[10], the weights were based on a response behaviour modelled using a logistic regression. The predictor variables were a set of socio-demographic respondent and

[10] At sweep 1 there was no prior information about the respondents so the modelling was based on information from the Child Benefit records, such as age of claimant, sex of claimant, number of children in the household and method of benefit payment. The other variables were Scottish index of multiple deprivation (quintiles), population density measured by the number of persons in private households per hectare and ONS urban rural indicator.

household characteristics collected from the previous sweeps. Non-response was associated with the following characteristics in all sweeps: renting the property, not working, being a younger mother (under the age of 20) and living in the 20% of the most deprived Data Zones. The predicted probability of response was then inversed to create the non-response weights. Hence, respondents who had a low predicted probability are allocated a larger weight, increasing their representation in the sample.

The final sweep weight is the product of the sweep's non-response weight and the previous sweep's interview weight. For each cohort the final weights were scaled to the responding sweep sample size to make the weighted sample size match the unweighted sample size.

A.4 Attrition among mixed families

As pointed out in Section A.2, the attrition rate among the *mixed* families is higher than the average attrition in the sample, particularly for the birth cohort. 71 *mixed* families drop out of the CC by wave 4 and 189 disappear from the BC1. This raises concerns for representativeness of the group, if the weighting applied in the study does not correct sufficiently for it.

Weights are created on a basis of logistic model detecting characteristics of respondents in the sample related to higher likelihood of attrition, which include lower household incomes, lone parent households, households with younger mothers and living in the more deprived areas. The weighting applied to the data will not work well for *mixed* families if their attrition is driven by different characteristics.

I compare the dropouts to the stayers among the *mixed* families to identify differences between them which may be related to attrition, focusing in particular on characteristics identified as correlated with attrition in the overall sample. I find that a higher proportion of dropouts have low household incomes and a higher proportion of respondents who drop out are young mothers. Those dropping out are also more likely to live in more deprived areas of Scotland. The differences within this group, however, are significantly smaller than for the group of native respondents. I also find no difference in % of lone parents among stayers and dropouts, which is identified as a determinant of attrition for the overall sample. Thus, the characteristics related to attrition are more pronounced for natives. Although the patterns are maintained for *mixed* families, no very clear selection emerges.

Results of a logistic regression of non-response on the family status (mixed vs. native) confirm that the *mixed* families are more likely to disappear from the study. To investigate whether different socio-economic characteristics trigger attrition of *mixed* and *native* families, I replicate the analysis undertaken by the authors of the data, to identify the characteristics correlated with attrition. I then repeat the same analysis on two subsets of data - for *mixed* and *native* families. I find that different characteristics matter to both groups, although they are all related to lower socio-economic outcomes of families and there is a degree of overlap in factors which matter. The elements also vary in importance - some factors are more influential for attrition among natives than among foreign-born. Overall the associations are weaker for *mixed* families suggesting that attrition is closer to random than for *native* families. Hence, the weights proposed in the study may not be most suitable for the purpose of my analysis.

The question is whether the weights matter at all then. I repeat all regressions presented in the paper on the unweighted data and find that the results remain unchanged, which is reassuring. All results for this analysis can be provided upon request.

A.5 Variables of interest

Table 17: Main variables used in the analysis

Child indicators	
Age	
Gender	
Language spoken at home	The data set contains information whether in the language spoken in the house is: English, English and another language, a language different from English only.
Family socio-economics	
NS-SEC category	National Statistics Socio-economic Classification classifies groups on basis of employment relations, career prospects, autonomy, mode of payment
NS-SEC category	and period of notice. The GUS dataset includes a five category system.
Highest education level	
Equivalised household annual income	Income data collected during the interview are adjusted to reflect the household composition.
Scottish Index of Multiple Deprivation	SIMD identifies areas of concentration of multiple deprivation across Scotland, based on indicatiors such as Current Income, Employment, Health, Education Skills and Training, Geographic Access to Services, Housing and Crime. The zones are then ranked from most deprived to least deprived.
Lone parents	Whether a parent is a lone parent.
Study child's birth order	
Number of siblings	
Parental investment measures	
Activities at home	
how often takes a child to park	
how often takes a child to friends who have children	
how often read books	
painting or drawing last week	
nursery rhymes last week	
letters and shapes practiced last week	
frequency played computer last week	
hours of TV watched	
Physical activity	
child rode a bicycle last week	
child kicked a ball last week	
child danced last week	
child ran/jumped last week	
child swam last week	
Outings	
Has been to library since last year	
Has been to a concert/play since last year	
Has been to a swimming pool since last year	
Has been to a museum, gallery etc since last year	
Has been to a zoo, aquarium etc since last year	
Has been to a cinema since last year	
Has been to a sports event since last year	
Disciplining techniques	
rewards/stickers	Respondents were asked whether they used the disciplining techniques with the child (yes/no question).
naughty step	
time out	
Child outcomes	
BAS Picture Recognition	The British Ability Scales aims to measure cognitive abilities and educational achievements of children. The Vocabulary Naming Exercise assessed a spoken vocabulary and may reflect expressive language ability, vocabulary knowledge of nouns, ability to attach labels to pictures, general knowledge and level of language stimulation. The Picture Similarities Exercise measures the reasoning ability of children. It may reflect non-verbal problem solving, visual perception and analysis, ability to attach meaning to pictures, use of verbal mediation and general knowledge. The scores are expressed as percentiles.
BAS Vocabulary Naming	
Subjective speech assessment:	
Can child be understood by respondent?	The assessment is based on asking the respondent to score the level of understanding using the following scale: mostly, sometimes and rarely. It is measured at ages 2,3,4 and 5.
Can child be understood by family?	
Can child be understood by strangers?	
Communication and Symbolic Behaviour Scales Score	This is an outcome derived only for the Birth Cohort at age 2. It is a measure of non-cognitive skills, assessing the child's communication, emotional development, understanding and interaction with peers. The composite score ranges from 0 to 57 and higher scores indicate better performance.
Strength and Difficulties Questionnaire Score	It is a score based on a behavioural screening questionnaire and measures: emotional symptoms, conduct problems, hyperactivity, inattention, peer relationship problems and pro-social behaviour. The overall score is a sum of all aspects, except the pro-social score, and the higher score indicates greater evidence of difficulties.

Table 18: Questions asked to compile SDQ and CSBS scores

Panel A: SDQ score	
Total score = emotional symptoms + conduct problems + hyper-activity + peer problems	
Emotional symptoms score	
MSDQ03	X often complains of headaches, stomach-aches or sickness
MSDQ08	X has many worries, often seems worried
MSDQ13	X is often unhappy, down-hearted or tearful
MSDQ16	X is nervous or clingy in new situations, easily loses confidence
MSDQ24	X has many fears, is easily scared
Conduct problems score	
MSDQ05	X often has temper tantrums or hot tempers
MSDQ07	X is generally obedient, usually does what adults request
MSDQ12	X often fights with other children or bullies them
MSDQ18	X often lies or cheats
MSDQ22	X steals from home, school or elsewhere
Hyper-activity score	
MSDQ02	X is restless, overactive, cannot stay still for long
MSDQ10	X is constantly fidgeting or squirming
MSDQ15	X is easily distracted, concentration wanders
MSDQ21	X thinks things out before acting
MSDQ25	X sees tasks through to the end, good attention span
Peer problems score	
MSDQ06	X is rather solitary, tends to play alone
MSDQ11	X has at least one good friend
MSDQ14	X is generally liked by other children
MSDQ19	X is picked on or bullied by other children
MSDQ23	X gets on better with adults than with other children
Pro-social score	
MSDQ01	X is considerate of other people's feelings
MSDQ04	X shares readily with other children (treats, toys, pencils etc.)
MSDQ09	X is helpful if someone is hurt, upset or feeling ill
MSDQ17	X is kind to younger children
MSDQ20	X often volunteers to help others (parents, teachers, other children)
Note: all questions in SDQ had the following possible answers: 1) not true, 2) somewhat true, 3) certainly true	
Panel B: CSBS score	
Social composite = emotion and eye gaze + communication + gestures	
Cluster 1: Emotion and eye gaze	
MCSBS01	Do you know when X is happy and when X is upset?
MCSBS02	When X plays with toys, does he look at you to see if you are watching?
MCSBS03	Does X smile or laugh while looking at you?
MCSBS04	When you look at and point to a toy across the room, does X look at it?
Cluster 2: Communication	
MCSBS05	Does X let you know that he needs help or wants an object out of reach?
MCSBS06	When you are not paying attention to X, does he try to get your attention?
MCSBS07	Does X do things just to get you to laugh?
MCSBS08	Does X try to get you to notice interesting objects - just to get you to look at the objects, not to get you to do anything with them?
Cluster 3: Gestures	
MCSBS09	Does X pick up objects and give them to you?
MCSBS10	Does X show objects to you without giving you the object?
MCSBS11	Does X wave to greet people?
MCSBS12	Does X point to objects?
MCSBS13	Does X nod his head to indicate yes?
Speech composite = sounds + words	
Cluster 4: Sounds	
MCSBS14	Does X use sounds or words to get attention or help?
MCSBS15	Does X string sounds or words together such as uh oh, mama, gaga, bye
MCSBS16	About how many of these sounds does X use: ma, na, ba, da, ga, wa, la, ya, sa , sha?
Cluster 5: Words	
MCSBS17	About how many different words does X use so that you know what he means ?
MCSBS18	Does X put two words together (such as 'more biccies'; bye-bye)?
Symbolic composite = understanding of words + object use	
Cluster 6: Understanding	
MCSBS19	When you call X's name, does he respond by looking or turning toward you?
MCSBS20	About how many different words or phrases does X understand without showing or pointing?
Cluster 7: Object use	
MCSBS21	Does X show interest in playing with a variety of objects?
MCSBS22	About how many of the following objects does your child use appropriately: cup, bottle, bowl, spoon, comb or brush, toothbrush, washcloth, ball, toy vehicle, toy, telephone?
MCSBS23	About how many blocks (or rings) can X stack?
MCSBS24	Does your child pretend to play with toys (for example, feed a stuffed animal, put a doll to sleep, put an animal figure in a vehicle)?
Note: all questions in CSBS has the following possible answers: 1) not yet, 2) sometimes, 3)often	

A.6 Explanatory variables

Some variables in the data set which I use in regression analysis are specific to the Scottish data. I briefly discuss how they are created and what they reflect. The explanations come from Bradshaw, Tipping, et al. (2009) which provides an overview of all the variables in the data set.

The National Statistics Socio-economic Classification (NS-SEC) is a social classification system that classifies groups on basis of employment relations, including career prospects, autonomy, mode of payment and period of notice. In GUS the classification contains 5 employment categories: managerial and professional, intermediate, small employers and own account workers, lower supervisory and technical, and semi-routine and routine occupations. The data set contains categorisations for the respondent, partner and the household as a whole. I use the household NS-SEC classification in the analysis. Further information on NS-SEC is available from the National Statistics website at:

http://www.statistics.gov.uk/methods_ quality/ns_sec/cat_subcat_class.asp

The Scottish Index of Multiple Deprivation (SIMD) identifies small area concentrations of multiple deprivation across Scotland based on seven individual domains of Current Income, Employment, Health, Education Skills and Training, Geographic Access to Services, Housing and new Crime. It is obtained at data zone level, ranking areas of median population size of 769, from the most deprived to the least deprived. In the dataset, the data zones are grouped into quintiles. Further details on SIMD can be found on the Scottish Government Website: http://www.scotland.gov.uk/Topics/Statistics/SIMD/Overview

The Equivalised Household Annual Income variable is a household income variable adjusted for a household's size and composition. Official income statistics use the 'Modified OECD' equivalence scale, in which an adult couple with no dependent children is taken as the benchmark with an equivalence scale of one and the scale is adjusted accordingly for other configurations within the household. The distribution of income for the population of the United Kingdom as a whole is taken from the most recent available data from the Family Resources Survey. The data and methodology are the same as those used by the Government in its annual Households Below Average Income publication.

B Data summary

Table 19: Mixed families and language spoken in the sample

Panel A: Percent of children with a parent born outside of the UK			
	Birth cohort and child cohort		
	foreign mother	foreign father	either parent foreign
sweep1	6%	5%	9%
sweep 2	6%	5%	5%
sweep 3	6%	5%	5%
sweep 4	6%	5%	5%
	Birth cohort only		
sweep 5	7%	6%	6%
sweep 6	6%	5%	5%

Panel B: Language spoken at home			
	only English	English and other	other only
overall	94%	5%	1%
mixed family	51%	40%	9%
native family	99%	1%	0%

C Further info on child outcomes

Table 20: Differences in development milestones measures for birth and child cohort

Birth cohort - baby development measures at age 1 (Δ%)		Child cohort - toddler development measures at age 3 (Δ%)	
Has child smiled at the respondent when he/she smiled at him?		**Is child able to walk on the level without difficulties?**	
often	0.03	yes	0.29
once or twice	-0.01	**Can child walk up steps like an adult, one foot on each step?**	
not yet	-0.02	Yes, alone	-1.9
Has child sat without being supported?		Yes with help	1.29
often	1.45	No	0.61
once or twice	-0.86	**Can child balance on one foot for at least four seconds?**	
not yet	-0.59	yes	-2.08
Has child stood up while holding onto something such as furniture?		**Can child hop at least twice on one foot?**	
often	2.57	yes	0.83
once or twice	-2.25	**Can child throw a ball?**	
not yet	-0.33	yes	0.11
Has child put his/her hands together?		**Can child grasp and handle small objects?**	
often	-0.81	yes	-0.34
once or twice	-0.51	**Can child undo big buttons?**	
not yet	1.32	yes	-3.07
Has child grabbed objects using his/her whole hand?		**Can child draw a circle?**	
often	-0.38	yes	0.44
once or twice	0.52	**Can child hold a pencil and scribble?**	
not yet	-0.15	yes	-0.24
Has child picked up a small object using forefinger and thumb only?		**Can child copy a square?**	
often	-1.71	yes	-0.23
once or twice	1.49	**Can child drink from a cup?**	
not yet	0.22	yes	-0.08
Has child passed a toy back and forth from one hand to another?		**Can child brush his/her teeth without help at least some of the time?**	
often	-1.23	yes	-3.44
once or twice	1.11	**Can child put on a t-shirt by him/herself?**	
not yet	0.12	yes	-3.47
Has child walked a few steps on his/her own?		**Can child get dressed without any help?**	
often	-1.37	yes	-3.31
once or twice	-0.69		
not yet	2.06		
Has child reached out and given you a toy or some other object that he/she is holding?			
often	-5.52		
once or twice	2.27		
not yet	3.26		
Has child waved bye-bye on his/her own when someone leaves?			
often	-12.13		
once or twice	2.7		
not yet	9.43		
Has child extended his/her arms to show he/she wants to be picked up?			
often	-1.1		
once or twice	0.72		
not yet	0.38		
Has child nodded his/her head for yes?			
often	-0.49		
once or twice	-1.05		
not yet	1.56		

Note: Here Δ is calculated as a difference between outcomes of children in mixed vs native families.

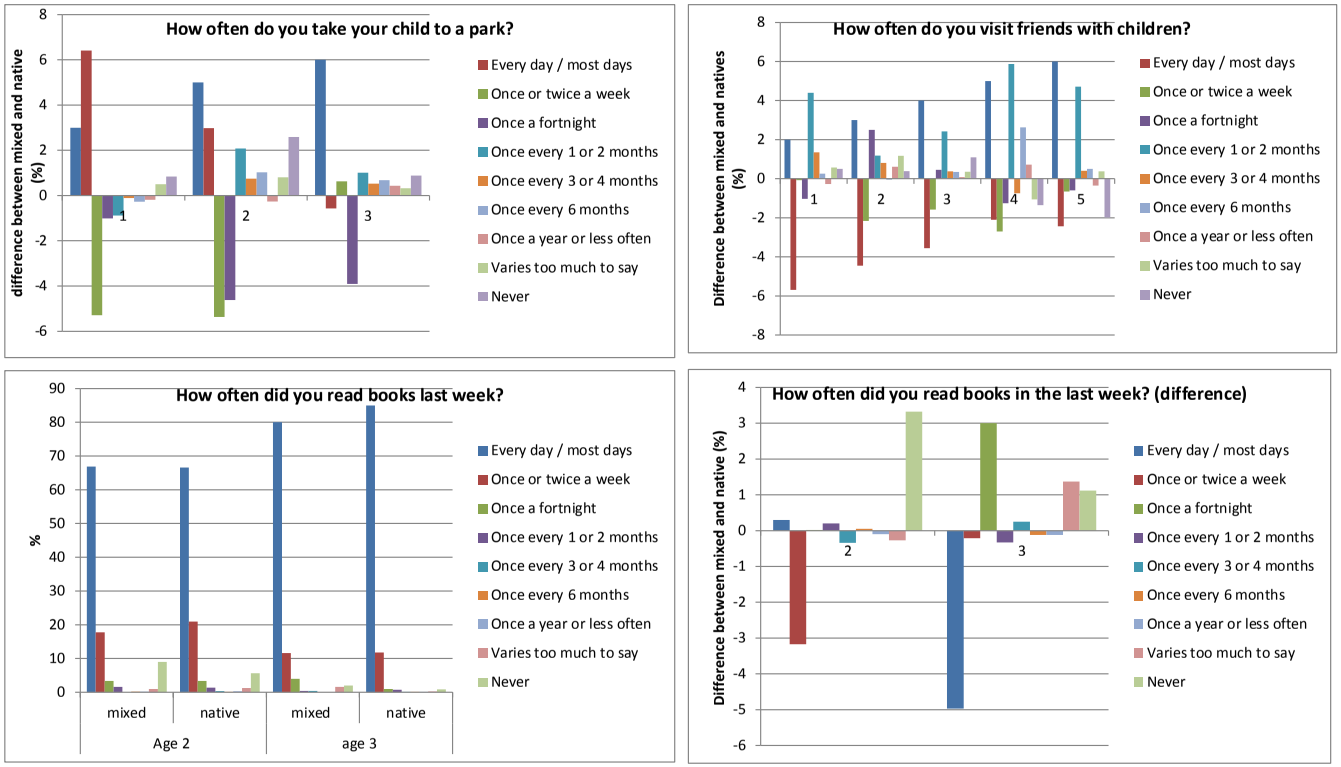


Figure 2: Raw differences in activities families engage in

C.1 Plots of raw differences in investments and outcomes

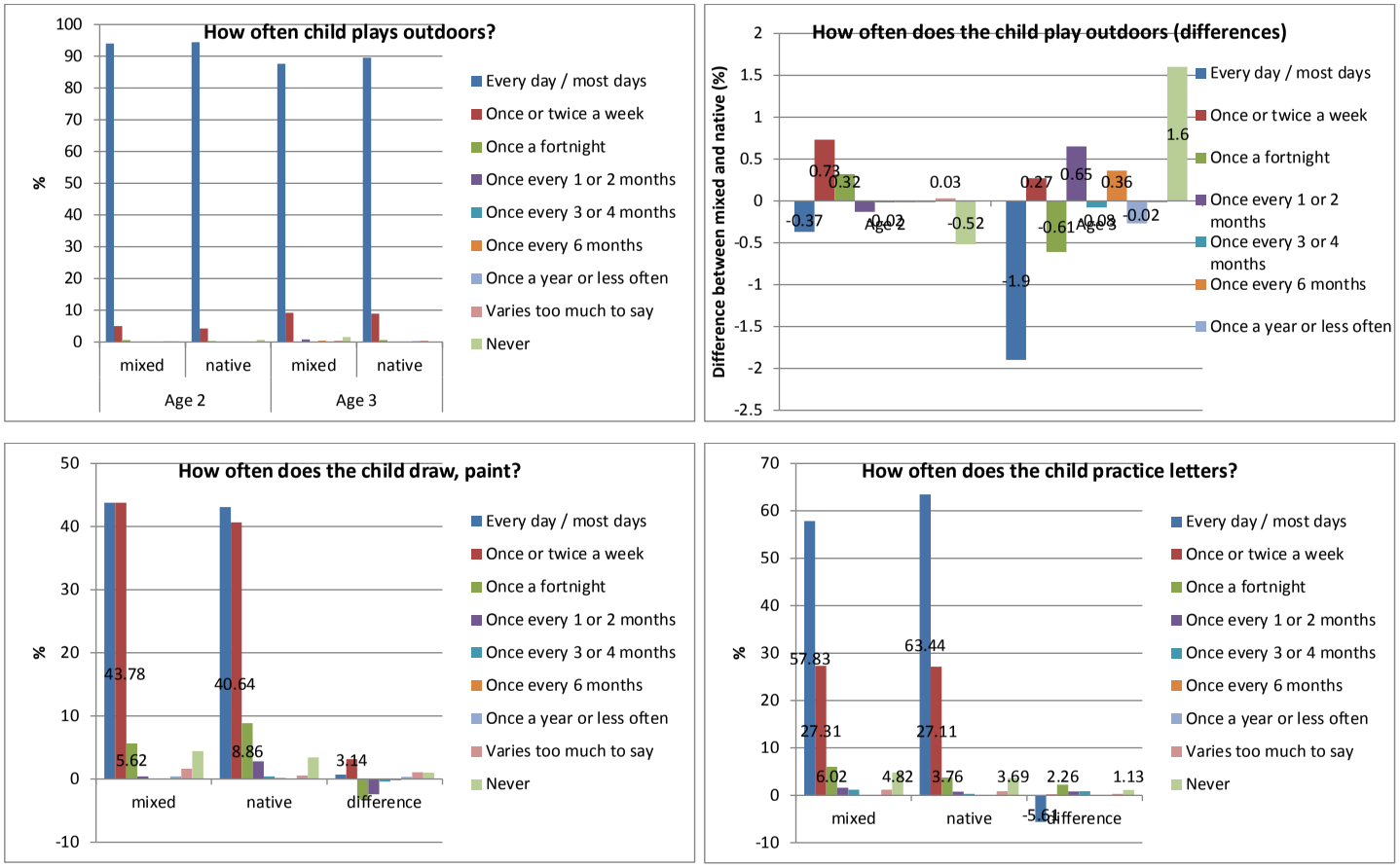


Figure 4: Raw differences in activities families engage in continued

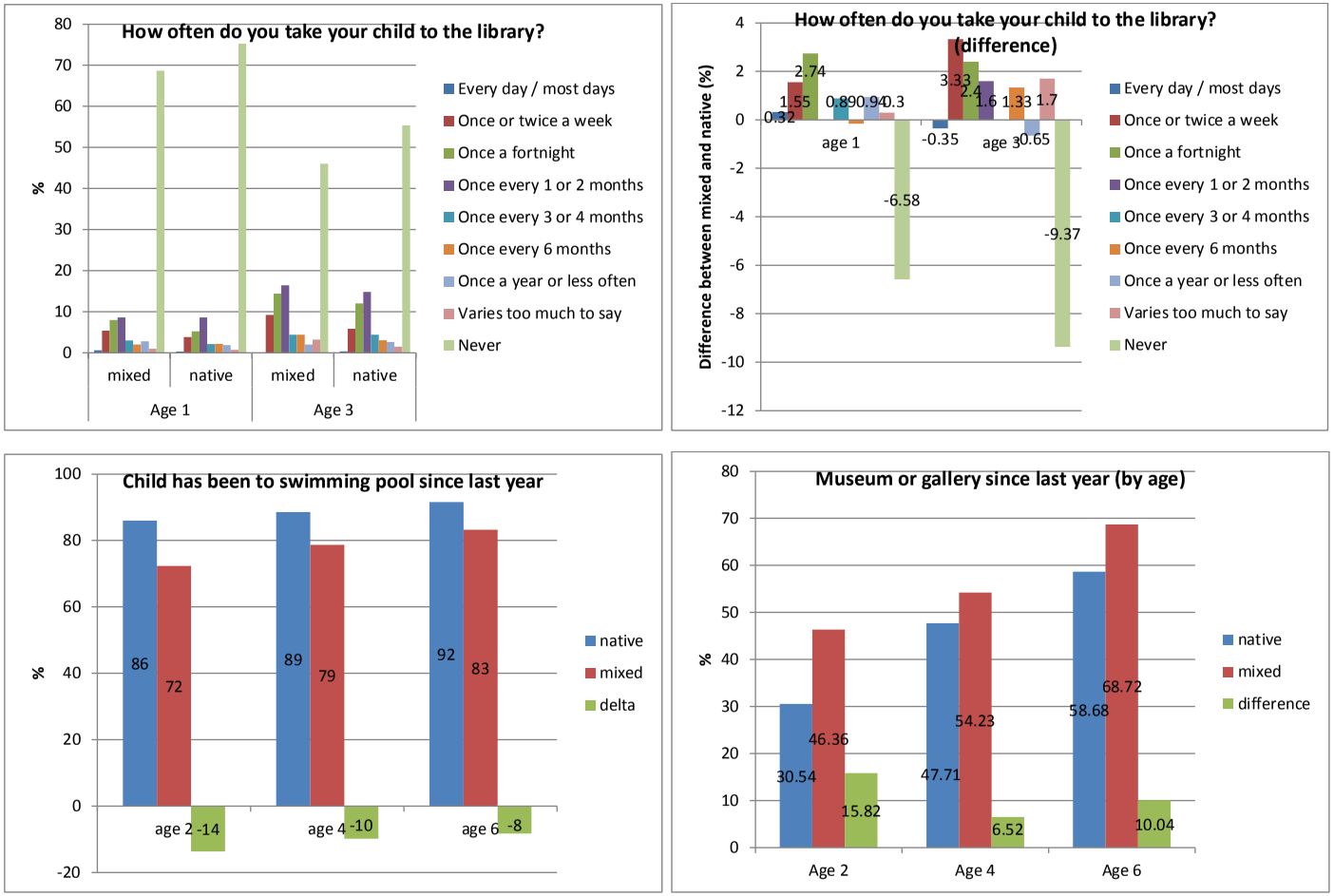
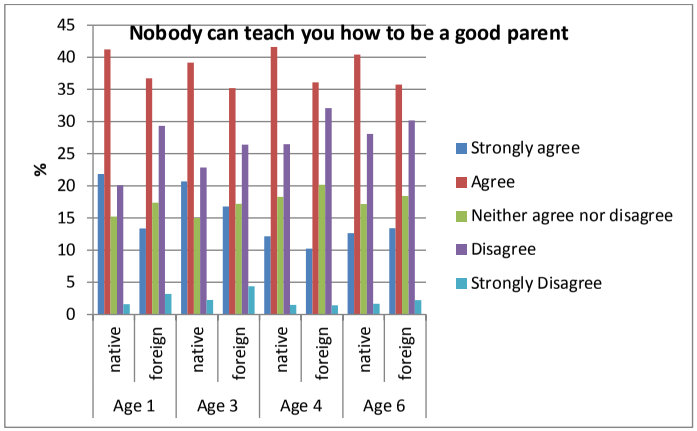
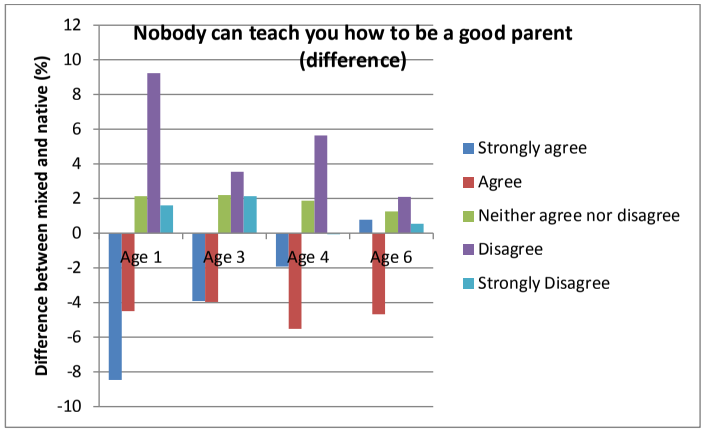
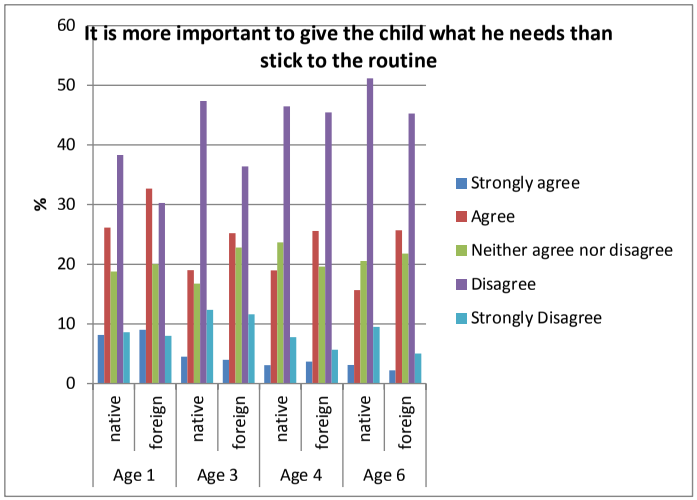


Figure 6: Raw differences in events attended by families







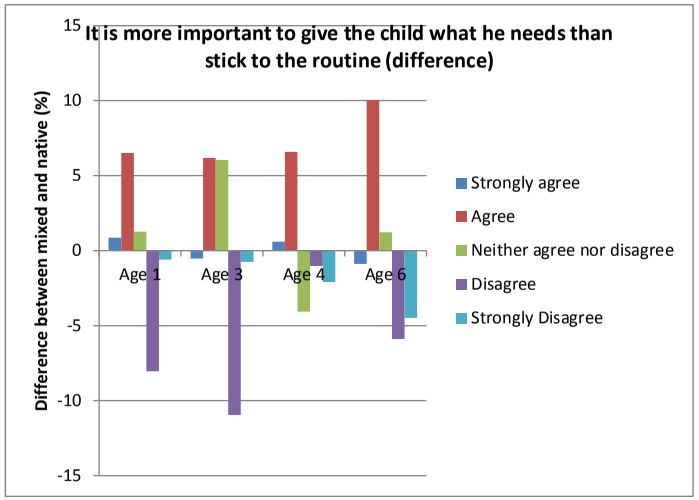


Figure 8: Raw differences in respondents' opinions

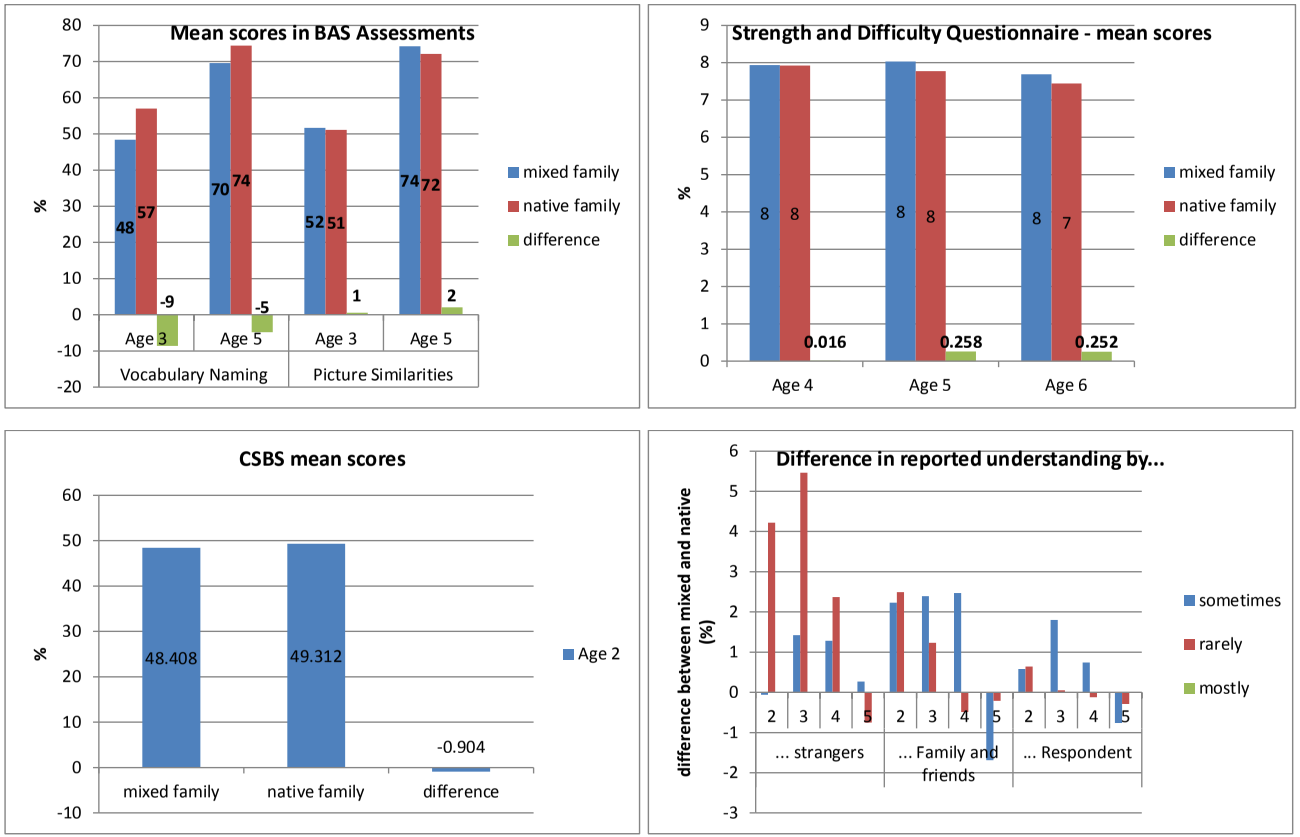


Figure 10: Raw differences in children's outcomes

D Further regression results

Table 21: Regression results for other speech indicators not influenced by language or mixed family status

Panel A: Can child be understood by family and friends?													
	(1) ordered probit	(2) ordered probit	(3) ordered probit	(4) ordered probit	(5) ordered probit	(6) ordered probit	(7) ordered probit	(8) ordered probit	(9) ordered probit	(10) ordered probit	(11) ordered probit	(12) ordered probit	(13) ordered probit
mixed	.096**	.100**	.104*	.012	.003	.006							-.028
	(.044)	(.051)	(.054)	(.082)	(.082)	(.082)							(.116)
language							.151***	.128*	.097	.049	.018	.020	-.078
							(.056)	(.070)	(.070)	(.100)	(.100)	(.101)	(.231)
mixed*language													.153
													(.280)
Controls:													
household controls	no	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	yes
activities at home	no	no	yes	yes	yes	yes	no	no	yes	yes	yes	yes	yes
outings	no	no	no	yes	yes	yes	no	no	no	yes	yes	yes	yes
physical activity	no	no	no	no	yes	yes	no	no	no	no	yes	yes	yes
discipline	no	no	no	no	no	yes	no	no	no	no	no	yes	yes
N	7451	6791	6617	4058	4058	4058	7447	6790	6616	4061	4058	4058	4058
pseudo R-squared	.094	.120	.135	.135	.102	.103	.094	.120	.135	.094	.102	.103	.103
Panel B: Can child be understood by the respondent?													
	(1) ordered probit	(2) ordered probit	(3) ordered probit	(4) ordered probit	(5) ordered probit	(6) ordered probit	(7) ordered probit	(8) ordered probit	(9) ordered probit	(10) ordered probit	(11) ordered probit		
mixed	.068	.053	.050	-.053	-.054						-.007		
	(.057)	(.066)	(.072)	(.129)	(.133)						(.177)		
language						.201***	.192**	.178*	-.003	-.005	.246		
						(.076)	(.094)	(.092)	(.168)	(.175)	(.370)		
mixed*language											-.336		
											(.449)		
Controls:													
household controls	no	yes	yes	yes	yes	no	yes	yes	yes	yes	yes		
activities at home	no	no	yes	yes	yes	no	no	yes	yes	yes	yes		
outings	no	no	no	yes	yes	no	no	no	yes	yes	yes		
physical activity	no	no	no	yes	yes	no	no	no	yes	yes	yes		
discipline	no	no	no	no	yes	no	no	no	no	yes	yes		
N	7451	6791	6617	4058	4058	7447	6790	6616	4058	4058	4058		
pseudo R-squared	.099	.114	.150	.113	.118	.100	.114	.150	.113	.118	.119		

Note: All regressions include gender, sweep and the cohort dummy as controls. Household controls include: number of siblings, parental education, NS-SEC household classification, equivalised income, geographical area and whether one is a single parent.Activities at home are frequency of painting/drawing, practicing rhymes, practicing letters, use of computer and TV. Outings include visits to many attractions, including museums, library, etc., physical activity includes swimming, playing in the park, running, etc. And disciplining techniques include time out and naughty step. The independent variable mixed is a dummy equal to 1 if at least one of child's parents was born outside of the UK and zero otherwise. Language is a dummy variable equal to 1 if child speaks English and another language or just another language at home. Error clustered at individual level in parentheses.
Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1

Table 22: Regression results for other measures not influenced by language or mixed family status

Panel A: CSBS score - clusters

	Emotion and eye gaze	Communication	Gestures	Sounds	Words	Understanding of words	Use of objects
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	ordered probit	ordered probit	ordered probit	ordered probit	ordered probit	ordered probit	ordered probit
mixed	-.015	.021	-.234***	-.116	-.128	-.181***	-.311***
	(.056)	(.081)	(.086)	(.080)	(.091)	(.060)	(.074)
all controls	yes	yes	yes	yes	yes	yes	yes

CSBS - individual elements of the affected clusters

Gestures cluster

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	picks objects		*shows objects*		*waves*		*points to objects*		*nods*	
	OLS	ordered probit	OLS	ordered probit	OLS	ordered probit	OLS	ordered probit	OLS	ordered probit
mixed	-.028	-.172	-.053	-.118	-.040	-.109	-.059**	-0.302***	-.041	-.062
	(.018)	(.107)	(.034)	(.076)	(.031)	(.093)	(.023)	(.107)	(.043)	(.079)
all controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

	Understanding of words						**Use of objects**					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	responds to name		*no of words known*		*plays with variety of objects*		*appropriate use of objects*		*stacking blocks*		*pretend play*	
	OLS	ordered probit	OLS	ordered probit	OLS	ordered probit	OLS	ordered probit	OLS	ordered probit	OLS	ordered probit
mixed	-.012	-.077	-.168***	-.241***	-.024	-.205	-.119***	-.354***	-.056	-.126	-.109***	-.268***
	(.017)	(.115)	(.053)	(.077)	(.016)	(.137)	(.036)	(.089)	(.037)	(.081)	(.034)	(.079)
all controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

Panel B: SDQ - individual elements of the affected clusters

Peer problems score

	solitary		has a good friend		liked by other children		picked on		gets on better with adults	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	probit	OLS	probit	OLS	probit	OLS	probit	OLS	probit
mixed	.018	.055	.008	.113	.006*	.329	.016	.099	.022	.069
	(.022)	(.064)	(.008)	(.120)	(.003)	(.238)	(.015)	(.083)	(.022)	(.069)
all controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	probit	OLS	probit	OLS	probit	OLS	probit	OLS	probit
language	.056*	.161*	-.009	-.091	.007*	.405*	.045**	.230**	.108***	.323***
	(0.031)	(.085)	(.014)	(.137)	(.004)	(.326)	(.023)	(.104)	(.031)	(.088)
all controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

Other questions

	Considerate of people's feelings		shares with other children		often has tantrums		fights with other children		has many fears	
	OLS	probit	OLS	probit	OLS	probit	OLS	probit	OLS	probit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
mixed	-.032***	-.309***	-.027**	-.309***	-.048**	-.138**	.021	.128	.037*	.113*
	(.012)	(.095)	(.011)	(.100)	(.022)	(.062)	(.015)	(.083)	(.022)	(.065)
all controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
	OLS	probit	OLS	probit	OLS	probit	OLS	probit	OLS	probit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
language	-.043**	-.358***	-.032**	-.352***	-.061**	-.179**	.037*	.203*	.016	.051
	(.017)	(.120)	(.015)	(.129)	(.031)	(.085)	(.021)	(.107)	(.029)	(.089)
all controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes

Note: no other components of the overall SDQ score are influenced by language spoken at home or being in a mixed family.
All regressions include gender, sweep and the cohort dummy as controls. Household controls include: number of siblings, parental education, NS-SEC household classification, equivalised income, geographical area and whether one is a single parent.Activities at home are frequency of painting/drawing, practicing rhymes, practicing letters, use of computer and TV. Outings include visits to many attractions, including museums, library, etc., physical activity includes swimming, playing in the park, running, etc. And disciplining techniques include time out and naughty step. The independent variable mixed is a dummy equal to 1 if at least one of child's parents was born outside of the UK and zero otherwise.
Language is a dummy variable equal to 1 if child speaks English and another language or just another language at home. Error clustered at individual level in parentheses.
Significance levels: *** p < 0.01, ** p < 0.05, * p < 0.1

Table 23: Regression results for motor and physical development indicators

Panel A: Baby measures (birth cohort only, age 1)

	child smiles	sits	stands up	puts hands together	grabs objects	picks up objects	passes toys	walks	reaches out	waves	extends arms	nods
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	oprobit	oprobit	oprobit	oprobit	oprobit	oprobit	oprobit	oprobit	oprobit	oprobit	oprobit	oprobit
mixed	3.182***	.367	.011	-.032	-.150	-.202**	-.140	-.062	-.144**	-.274***	-.071	-.039
	(.137)	(.231)	(.075)	(.090)	(.216)	(.092)	(.110)	(.073)	(.068)	(.061)	(.077)	(.075)
all controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	4671	4671	4671	4669	4670	4649	4663	4671	4670	4669	4670	4666
pseudo-R-squared	.091	.056	.008	.013	.020	.016	.010	.008	.010	.025	.007	.014
language	-.319	.031	.042	-.175	-.201	-.340***	-.103	-.031	-.146	-.346***	.033	.105
	(.350)	(.230)	(.101)	(.112)	(.261)	(.113)	(.143)	(.096)	(.090)	(.083)	(.116)	(.094)
all controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	4671	4671	4671	4669	4670	4649	4663	4671	4670	4669	4670	4666
pseudo-R-squared	.087	.052	.008	.014	.020	.017	.009	.008	.009	.024	.007	.014

Note: The answers to baby measures are coded in the following way: 0 - no, 1-sometimes, 2-often. Hence, a positive coefficient indicates better performance

Panel B: Toddler measures (child cohort, age 3)

	walks up steps	balances on foot	hops on foot	throws a ball	grasps objects	undoes buttons	draws a circle	scribbles	copies a square	drinks from a cup	brushes teeth	puts a T-shirt on	dresses himself
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	probit	probit	probit	probit	probit	probit	probit	probit	probit	probit	probit	probit	probit
mixed	-.104	-.105	.029	.041	-.263	-.065	.022	-.309	-.002	-.063	-.323	-.060	-.064
	(.122)	(.073)	(.065)	(.311)	(.223)	(.065)	(.067)	(.193)	(.061)	(.156)	(.088)	(.063)	(.060)
all controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	6443	6170	6216	6444	6444	6246	6371	6444	5967	6443	6438	6411	6430
pseudo-R-squared	.002	.004	.009	.024	.039	(.031)	.030	.023	.008	.010	.012	.051	.074
language	-.160	-.139	-.052	-.428	-.306	-.038	.004	-.558***	-.019	-.076	-.515***	-.119	-.118
	(.151)	(.094)	(.084)	(.265)	(.268)	(.086)	(.087)	(.209)	(.080)	(.202)	(.103)	(.081)	(.078)
all controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	6442	6169	6215	6443	6443	6245	6370	6443	5966	6442	6437	6410	6429
pseudo-R-squared	.002	.004	.009	.171	.039	.030	.030	.034	.008	.010	.017	.051	.074

Note: Measures of toddler development were coded as dummy variables equal to 1 if a child can do a given thing.

Controls in the regressions include: child's gender, number of siblings, mother's education, household equivalised income, area of deprivation and whether family a full family.
Robust standard errors in parentheses, *** p < 0.01, ** p < 0.05, * p < 0.1

References

Aizer, A., and F. Cunha (2012): "The production of human capital: endowments, investments and productivity.," NBER Working Paper Series 18429.

Allen, W. (2013): "Geographical Distribution and Characteristics of International Migration Flows to Scotland," Discussion paper, The Migration Observatory, Oxford.

Apps, P., S. Mendolia, et al. (2012): "The impact of pre-school on adolescents' outcomes: evidence from a recent English cohort," IZA Discussion Paper 6971.

Bak, T., J. Nissan, et al. (2014): "Does bilingualism influence cognitive aging?," *Annals of Neurology*, 00, 1–5.

Baker, C. (1999): *Foundations of Bilingual Education and Bilingualism*. Multilingual Matters, Bristol.

Becker, B. (2010): "Who Profits Most from Early Parental Investments? The Effects of Activities Inside and Outside the Family on German and Turkish Children's Language Development," *Child indicators research*, 3, 29–46.

Behrman, J. R., J. Hoddinott, et al. (2006): "What Determines Adult Cognitive Skills? Impacts of Pre-Schooling, Schooling and Post-Schooling Experiences in Guatemala," *Journal of Economic Literature*.

Bialystok, E. (2009): "Bilingualism: the good, the bad and the indifferent," *Bilingualism: language and cognition*, 12, 3–11.

Bjorklund, A., and K. Salvanes (2010): "Education and Family Background: Mechanisms and Policies," IZA Discussion Paper 5002.

Black, S., and P. Devereux (2011): "Recent Developments in Intergenerational Mobility," in *Handbook of Labor Economics*. Elsevier.

Black, S., P. Devereux, et al. (2005): "Why the apple doesn't fall far: understanding intergenerational transmission of human capital," *American Economic Review*, 95, 437–449.

Boehlmark, A. (2008): "Age at immigration and school performance: A siblings analysis using Swedish register data," *Labour Economics*, 15, 1366–1387.

Borghans, L., A. Lee Duckworth, et al. (2008): "THE ECONOMICS AND PSYCHOLOGY OF PERSONALITY TRAITS," NBER Working Paper 13810.

Bradshaw, P., S. Tipping, et al. (2009): "Growing Up in Scotland, Sweep 5. User guide," Scottish Centre for Social Research.

Brooks-Gunn, J., and L. Markman (2005): "The Contribution of Parenting to Ethnic and Racial Gaps in School Readiness," *The Future of Children*, 15, 139–168.

Carneiro, P., C. Crawford, et al. (2007): "Impact of early cognitive and non-cognitive skills on later outcomes," Discussion paper.

Carneiro, P., J. J. Heckman, et al. (2011): "Estimating Marginal Returns to Education," *The American Economic Review*, 101, 2754–2781.

Chiswick, B., and C. Houseworth (2011): "Ethnics intermarriage among immigrants: human capital and assortative mating," *Review of Economics of the Household*, 9(2), 149–180.

Conti, G., J. Heckman, et al. (2011): "Early Endowments, Education and Health," Human Capital and Economic Opportunity: A Global Working Group, Working Paper No. 2011-001.

Cunha, F., J. Heckman, et al. (2010): "ESTIMATING THE TECHNOLOGY OF COGNITIVE AND NONCOGNITIVE SKILL FORMATION," NBER Working Papers 15664.

De Philippis, M. (2014): "Parents' Country of Origin and Intergenerational Mobility of School Performance.," unpublished.

Docquier, F., and A. Marfouk (2006): "International migration by educational attainment (1990-2000) - Release 1.1," in *International Migration, Remittances and Development*, ed. by C. Ozden, and M. Schiff. Palgrave Macmillan, New York.

Dronkers, J., and M. de Heus (2012): "The Educational Performance of Children of Immigrants in Sixteen OECD Countries," Discussion paper.

Duncan, B., and S. Trejo (2011): "Intermarriage and the intergenerational transmission of ethnic identity and human capital for Mexican Americans," *Journal of Labor Economics*, 29, 195–227.

Dustmann, C., T. Frattini, et al. (2010): "Ethnicity and second generation immigrants in Britain," Discussion paper.

——— (2012): "Educational achievement of second-generation immigrants: an international comparison," *Economic Policy*, 27, 143–185.

Dustmann, C., S. Machin, et al. (2010): "Ethnicity and educational achievement in compulsory schooling," *The Economic Journal*, 120, F272–F297.

Eirich, F. (2011): "Characteristics and intentions of immigrants to and emigrants from Scotland - review of existing evidence," Discussion paper, Scottish Government Social Research, Edinburgh.

Ermisch, J. (2008): "Origins of Social Immobility and Inequality: Parenting and Early Child Development," *Journal of the National Institute of Economic and Social Research*, 205, 62–71.

Feinstein, L. (2003): "Inequality in the Early Cognitive Development of British Children in the 1970 Cohort," *Economica*, 70, 73–97.

Fuller, B., M. Bridges, et al. (2009): "The Health and Cognitive Growth of Latino Toddlers: At Risk or Immigrant Paradox?," *Maternal and Child Health Journal*, 13, 755–768.

Glick, J. E., L. D. Hanish, et al. (2012): "Migration Timing and Parenting Practices: Contributions to Social Development in Preschoolers With Foreign-Born and Native-Born Mothers," *Child Development*, 83, 1527–1542.

Glick, J. E., and B. Hohmann-Marriott (2007): "Academic Performance of Young Children in Immigrant Families: The Significance of Race, Ethnicity, and National Origins," *International Migration Research*, 41, 371–402.

Greenwood, J., N. Guner, et al. (2014): "Marry Your Like: Assortative Mating and Income Inequality," Discussion paper.

Hanushek, E. A., and L. Woessmann (2009): "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation," IZA Discussion Paper 4575.

Hartas, D. (2011): "Families' social backgrounds matter: socio-economic factors, home learning and young children's language, literacy and social outcomes," *British Educational Research Journal*, 37, 893–914.

Heckman, J., and Y. Rubinstein (2001): "The importance of noncognitive skills: lessons from the GED testing program," *The American Economic Review*, 91, 145–149.

Heckman, J. J., and G. Conti (2012): "The economics of child well-being," NBER Working Paper 18466.

HM Revenue and Customs (2012): "Child Benefit, Child Tax Credit and Working Tax Credit Take Up Rates 2010-11," Discussion paper.

Holmlund, H., M. Lindahl, et al. (2011): "The causal effect of parents' schooling on children's schooling: a comparison of estimation methods.," *Journal of Economic Literature*, 49, 615–651.

Kaushanskaya, M., and V. Marian (2007): "Bilingual Language Processing and Interference in Bilinguals: Evidence From Eye Tracking and Picture Naming," *Language Learning*, 57, 119–163.

Keane, M., and M. Fiorini (2012): "How the Allocation of Children's Time Affects Cognitive and Non-Cognitive Development," Discussion paper, University of Oxford, Department of Economics.

Kornder, N., and J. Dronkers (2012): "Do Migrant Girls Always Perform Better? Differences between the Reading and Math Scores of 15-Year-Old Daughters and Sons of Migrants in PISA 2009 and Variations by Region of Origin and Country of Destination," Discussion paper.

Lanzieri, G. (2012): "Mixed marriages in Europe, 1990-2010," in *Cross-Border Marriage: Global Trends and Diversity*, ed. by D.-S. Kim, pp. 81–121. Korea Institute for Health and Social Affairs (KIHASA).

Lichter, D., and Z. Qian (2001): "Measuring marital assimilation: intermarriage among natives and immigrants.," *Social Science Research*, 30, 289–312.

Meier Jaeger, M. (2011): "Does Cultural Capital Really Affect Academic Achievement? New Evidence from Combined Sibling and Panel Data," *Sociology of education*, 84, 281–298.

Melhuish, E., K. Sylva, et al. (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, 157–188.

Meng, X., and R. Gregory (2005): "Intermarriage and the economic assimilation of immigrants.," *Journal of Labor Economics*, 23, 135–176.

National Records of Scotland (2013a): "2011 Census: Detailed characteristics on Ethnicity, Identity, Language and Religion in Scotland - Release 3A," Discussion paper.

——— (2013b): "2011 Census: Key Results on Population, Ethnicity, Identity, Language, Religion, Health, Housing and Accommodation in Scotland - Release 2A," Discussion paper.

Nordin, M., and D.-O. Rooth (2007): "The Income Gap Between Natives and Second Generation Immigrants in Sweden: Is Skill the Explanation?," Discussion paper.

Nores, M., and S. Barnett (2010): "Benefits of early childhood interventions across the world: (Under)Investing in the very young," *Economics of Education Review*, 29, 271–282.

Oller, K., and R. Eilers (2002): *Language and Literacy in Bilingual Children*. Multilingual Matters.

Ozden, C., C. Parsons, et al. (2011): "Where on Earth is Every-body? The evolution of global bilateral migration 1960-2000.," *World Bank Economic Review*, 25(1), 12–56.

Portocarrero, J., R. Burright, et al. (2007): "Vocabulary and verbal fluency of bilingual and monolingual college students," *Archives of Clinical Neuropsychology*, 22, 415–422.

Reardon, S., and C. Galindo (2009): "The Hispanic-White Achievement Gap in Math and Reading in the elementary grades," *American Educational Research Journal*, 46.

Rienzo, C. (2013): "Characteristics and Outcomes of Migrants in the UK Labour Market," Discussion paper, The Migration Observatory, Oxford.

Rosenthal, A., K. Baker, et al. (1983): "The effect of language background on achievement level and learning among elementary school students," *Sociology of Education*, 56, 157–169.

Sandefur, G., and T. McKinnell (1986): "American Indian intermarriage," *Social Science Research*, 347, 347–348.

Vargas-Silva, C. (2013a): "Migrants in Scotland: An Overview," Discussion paper, The Migration Observatory, Oxford.

——— (2013b): "Migrants in the Scottish Labour Market," Discussion paper, The Migration Observatory, Oxford.

Wetherby, A., and B. Prizant (2001): "Communication and Symbolic Behavior Scales Developmental Profile," Paul H. Brookes Publishing Co.