

Cobb-Clark, Deborah A.; Harmon, Colm; Staneva, Anita

Article

The bilingual gap in children's language, emotional, and pro-social development

IZA Journal of Labor Economics

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Cobb-Clark, Deborah A.; Harmon, Colm; Staneva, Anita (2021) : The bilingual gap in children's language, emotional, and pro-social development, IZA Journal of Labor Economics, ISSN 2193-8997, Sciendo, Warsaw, Vol. 10, Iss. 1, pp. 1-41, <https://doi.org/10.2478/izajole-2021-0001>

This Version is available at:

<https://hdl.handle.net/10419/298390>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Deborah A. Cobb-Clark^{1,2}, Colm Harmon^{1,2,3} and Anita Staneva^{1,4*}

The bilingual gap in children's language, emotional, and pro-social development

Abstract

In this paper we examine whether – conditional on other child endowments and family inputs – bilingual children achieve different language, emotional, and pro-social developmental outcomes. Our data, which allow us to analyze children's development in a dynamic framework, are extracted from the UK Millennium Cohort Study (MCS). We model the development production functions for bilingual children using cumulative value-added specifications, which account for parental investments and children's own ability. Analysis based on child age confirms that bilingual children initially have worse language skills than their monolingual peers. The commencement of schooling appears to attenuate these differences, and by age seven, bilingual children have a developmental advantage. We find evidence of a positive relationship between bilingualism and some aspects of emotional development, and it is mainly boys who appear to benefit from their bilingual background.

Current version: November 23, 2020
Keywords: cognitive and non-cognitive skills, production function, value-added model, cohort studies
JEL-codes: I20; J24; D10
Corresponding author: Anita Staneva
a.staneva@griffith.edu.au

1 University of Sydney, School of Economics, Faculty of Art and Social Science, Australia

2 Institute for the Study of Labor Economics (IZA), Bonn

3 University of Edinburgh, Old College, Scotland

4 Griffith University, Griffith Business School, Department of Accounting, Finance and Economics, Australia

1 Introduction

Schools, especially in major urban cities, are used by a large number of children for whom English is a second language. These children speak numerous home languages and often constitute the majority of children in a single classroom (Bialystok, 2006). In the United States, for example, more than 4.9 million English learners were enrolled in public elementary and secondary schools during the 2013–14 school year, representing just over 10% of the total student population (National Center for Education Statistics, 2016), while in the Canadian province of Ontario over 25% of students are identified as English language learners (Ontario Education Department, 2007). Similarly, the UK Department of Education estimates that there were more than a million children in the 5–18 age group who were enrolled in UK schools, who collectively speak more than 360 different languages (Department for Education, 2013).

The rise in global migration has left these countries grappling with the challenges of educating and integrating bilingual children. For instance, the cost of providing English language support to 50 pupils has been estimated by Edinburgh City Council to be approximately £ 33,000 per year, and this sum is driven largely by the cost of employing an English language support teacher (Heather Rolfe and Hilary Metcalf, 2009). At the same time, there is a debate going on, which is largely centered around the economic merits of speaking a second language in the context of globalization – directly addressing the danger of a country that relies on the primacy of, say, English (see Nuffield Foundation (2002) for one example of the debate in relation to the UK). The wider question is, to what extent bilingualism matters in the early phase of a child’s development, or whether bilingual children perform better than English-only students (Bialystok et al., 2012; Hoff, 2013).

Studies can be cited to support either side of the debate. On the one hand, dual language learning in early childhood places additional burdens on language development in comparison to the acquisition of a single language (Hoff, 2013). On the other hand, there is some evidence in the neuroscience and developmental psychology literature which indicates that bilingual children develop enhanced executive control – i.e., an ability to effectively manage cognitive processes such as problem solving, memory, and thought – resulting in a “bilingual advantage” (Bialystok et al., 2012). Similarly, parents’ language choices may influence the way children express their emotions; this affects their emotional development. Shifting from one language to another, for example, helps children to regulate their emotional response by using a less emotional non-native language (Chen et al., 2012). Bilingual children tend to exhibit fewer behavioral problems than monolingual English and Spanish speakers (see Han, 2010 and Han and Huang, 2010) and have stronger family cohesion and higher self-esteem (Portes and Hao, 2002).

Motivated by the importance of early life experience in driving children’s life chances, this paper attempts to link these diverse strands of the literature by integrating the role of language in children’s skill formation. Building on the theoretical analyses of Todd and Wolpin (2003), Cunha et al. (2006), and Cunha et al. (2010), we examine whether bilingual children have different development outcomes on the basis of “conditional on child endowments” and other family inputs. More specifically, we estimate a skills production function for children born in the UK to at least one migrant parent. This focus on the early development of the native-born children of migrant parents allows us to avoid any potential confounding factors associated with the time which has elapsed since migration. It also provides the opportunity to

test whether, given the same educational environment, bilingual children have different developmental outcomes as compared to their monolingual peers. Finally, the specific challenges we identify and the policy solutions which consequently need to be implemented should be relevant to this particular group of children, namely UK born children who are born to at least one migrant parent.¹

We advance the literature available on the subject in several important ways. First, we explore the data available from the UK Millennium Cohort Study (MCS), which allows us to analyze children's language, emotional, and pro-social development in depth and within a dynamic framework. The strength of these data relative to other available data is that they provide unusually detailed longitudinal information on children's home environment. We estimate the bilingual gap in a child's language, emotional, and pro-social behavior as a cumulative process that depends on current and past endowments of cognitive and non-cognitive capacity, as well as on the history of family investments. Previous studies that have relied on cross-sectional data are only able to examine the contemporaneous relationship between bilingualism and child development. However, it is important to study this effect over time. Researchers have argued that although it takes several years of exposure to English for immigrant children to approach native-like fluency in conversational skills, it takes five to nine years to become proficient in academic English (see Collier, 1989 and August and Hakuta, 1997). Our examination of the bilingual gap in the acquisition of language, emotional, and pro-social skills between the ages of three and fourteen, complements existing studies that focus solely on the link between bilingualism and developmental skills which are measured at specific ages. To our knowledge, this study is the first to identify the evolution of bilingual effects across language, emotional, and pro-social behavior assessments from an early age to adolescence.

Second, studying the dynamic consequences of the language which is spoken at home on socio-emotional development is important. It has been shown that behavioral and emotional difficulties not only reflect impaired mental health and high probability of school dropout (Raver, 2003), but are also linked to adult outcomes, including educational attainment, earnings, and employment (see Najman et al., 2004 and Jones et al., 2015). Pro-sociality is another important aspect of human personality that affects a wide range of economic decisions, with both low and high pro-social behavior conferring the risk for adverse behavior and mental health difficulties in children (see Nantel-Vivier et al., 2014).² Engagement in pro-social activities requires a certain level of language competency since it is largely linguistically rooted.

Third, we explore a range of parental behaviors and home environment activities that can potentially drive the estimated language effects. We hypothesize that potential differences in children's language and pro-social behavior could arise from differences in maternal behavior and mother-child interactions. Finally, a potential problem when estimating the skill production function lies in accounting for unobserved heterogeneity that could confound the

1 Some elements in the literature do not distinguish between particular groups of migrant children (see for instance Heckmann (2008), Brind et al. (2008), and Eurodiaconia (2016)). In this way, the authors adopt a comprehensive approach which encompasses a wide range of issues that are relevant to migrant children, and policy solutions are not clearly differentiated.

2 Low pro-sociality may indicate a lack of empathy, thereby increasing the likelihood of disruptive, harmful behaviors towards others (Hastings et al., 2000). High pro-sociality is associated with high internalizing (depressive) symptoms, particularly in girls Keenan and Hipwell (2005).

relationship. We improve upon the existing evidence by utilizing value-added and dynamic panel data models. While we cannot interpret our results as being causal, the inclusion of past endowments and an estimation of the models with a comprehensive set of controls should largely reduce the bias arising from unobserved heterogeneity.³

The findings in this article have important implications for identifying children (at different ages) who are likely to lag behind and thus need additional support, and understanding the potential for policies, families, and schools to improve the learning outcomes for these children. Using an age-specific analysis, we find that children identified as bilingual have lower levels of language development in early childhood. The start of schooling appears to attenuate these differences and at age seven, once we account for past endowments, bilingual children have a language advantage. At ages 11 and 14, bilingual children are not significantly different relative to their monolingual peers; however, we find a language deficit for boys. Boys' underachievement at age 14 in verbal skills is consistent with some official figures. According to the 2011 National Literacy Trust data, 76% of boys achieved the expected level in English at Key Stage 3 (age 14) compared to 88% of girls (All-Party Parliamentary Literacy Group Commission et al., 2012). We show that there is evidence in maternal behavior – especially in the mothers' provision of cognitive stimulation and interactions at home – that can explain in part the relationship between the narrowing gap between bilingual and monolingual children and early language development. Mothers identified as bilingual are consistently more likely to spend their time with the child and to invest in home activities. In the case of emotional development, it is mainly boys that are found to benefit from their bilingual background, though the estimated bilingual effect varies based on age. Finally, we document a fair amount of heterogeneity in estimated effects along the mother's education. Highly educated mothers are more likely to speak English at home. Consequently, the language development of their children is found to be statistically equivalent to monolingual peers from the age of five onward. Thus, our findings may be particularly useful in development of intervention programs for immigrant families, in helping intervention staff to promote early language, emotional, and pro-social behavior, in helping immigrant families to remain aware of how the use of different languages can have an impact on their child's development.⁴

The rest of the paper is organized as follows. Section 2 discusses the previous literature and Section 3 presents the data, sample selection, and variables' construction. Section 4 outlines our conceptual framework and empirical specification. Section 5 discusses the results and Section 6 concludes the paper.

2 The Literature

While we know little about the relationship between bilingualism and emotional and pro-social development, there is an extensive literature in linguistic and developmental psychology on the impact of bilingualism on children's cognitive development. These studies are usually

3 Estimating the causal effect of bilingualism on children's outcomes is a challenging task. Our data do not provide us with any quasi-experimental variation which is suitable for causal identification. Indeed, it is hard to imagine an experiment that would lead to a random assignment of the language that children speak. Therefore, we regard our results as providing descriptive, non-causal evidence.

4 For example, there is evidence that school-based interventions promote pro-social behavior and therefore help in reducing aggressive behavior (Caprara et al., 2014).

motivated by the observation that bilingualism confuses children (Genesee, 1989), or that bilingual children are slower to develop language skills because their learning capacity is divided across acquiring multiple languages (Macnamara, 1967; Hoff, 2013). Studies of bilingual children which focus on verbal tests of intelligence appear to lend support to these concerns by concluding that, on average, bilingual children exhibit slower cognitive development (see Darcy, 1953; Hakuta, 1986; and Bhatia and Ritchie, 2008) and initially possess a smaller vocabulary in each of their languages (Oller and Eilers, 2002; Clifton-Sprigg, 2016). A lack of English proficiency is cited as a primary reason for first- and second-generation children's poor academic performance in elementary school (Rosenthal et al., 1983).

There is emerging evidence, however, of the importance of bilingualism for some forms of cognitive functioning. Bilingualism is associated with higher executive functioning and attention in children (Bialystok, 2001; Yang et al., 2011) and young adults (Costa et al., 2008), and seems to protect against cognitive decline in old age (Bialystok et al., 2012). Further, learning a foreign language may improve the development of analytical and communicative skills (Saiz and Zoido, 2005), while exposure to two languages at an early age is positively correlated with reading, phonological awareness, and language competence in both languages (Kovelman et al., 2008).

Moreover, dual language competency is also positively related to the emotional and behavioral well-being and school-functioning of children born to immigrant parents (Collins et al., 2011). Han (2010) and Leyendecker et al. (2014) find that bilingualism encourages and facilitates children's communication and that this has an impact on their socio-emotional and behavioral skills. At school, young children from immigrant families have been found to have particular strengths in the areas of social skills and behavior (Crosnoe, 2007; De Feyter and Winsler, 2009). Some other studies, however, contradict these findings. For example, Spomer and Cowen (2001) compare mental health referrals for children who were English language learners and native learners, and find that the English language learners were rated higher in shyness/anxiety, and lower in peer social skills and assertiveness, by their teachers.

Most of these studies do not provide causal evidence, since exogenous variation in bilingualism is rarely available. Exceptions include a few cohort and randomized control studies (see Slavin et al. (2011) and Chin et al. (2013)). However, these studies generally use data from outside the UK and do not examine the dynamic relationship aspect. Clifton-Sprigg (2016) examines the early age performance gap which prevails in bilingual children, using longitudinal data for Scotland via instrumental variable strategy. The authors find that children perform comparably on an array of measures, including cognitive (picture similarities), non-cognitive (strengths and difficulties questionnaire), and motor development. Where differences in performance do emerge (examples for these differences being vocabulary naming or speech assessment), the outcomes are likely to be related to speech and linguistic skills. The authors highlight that bilingual families are a heterogeneous group and that children with two foreign-born parents are at a particular disadvantage at this early age. Slavin et al. (2011) randomly assign kindergartners who have limited proficiency in English to either bilingual education or structured English "immersion" programs across six schools. They find no statistically significant difference in standardized test scores in English by fourth grade (i.e., by age nine or ten). Although students undergoing bilingual education initially had weaker English skills than students of the immersion programs, their later skills did not differ significantly. Chin et

al. (2013) exploit policy changes which govern the provision of bilingual education programs in Texas, where school districts were required to offer bilingual education when the enrollment of students who have limited English proficiency in a particular elementary grade level reaches a specific threshold. Using this discontinuity as an instrument for district bilingual education provision, they find that bilingual education programs do not significantly impact the standardized test scores of students with Spanish as their home language.

A number of studies use MCS data (as we do here) to examine the importance of home environment and time investment in shaping early child outcomes. Ermisch and Pronzato (2008) find that much of the difference in child development at age three can be explained by parenting style and educational activities. Similarly, Dickerson and Popli (2016) find that the cognitive development test scores at age seven are lower for children who live persistently below the poverty line throughout their early years, as compared to children who have never experienced poverty. Forrest et al. (2018) use MCS data to examine the impact of early language difficulty and peer problems on later emotional outcomes, and find that children with language difficulties are at an increased risk for social and emotional problems. However, the studies utilizing UK data do not typically account for unobservable child heterogeneity or for the persistence in language and emotional development over time. Some exceptions are Del Bono et al. (2016), who estimate the relationship between maternal time inputs and early child development using a specification in which lagged inputs and past test scores are controlled for, and (Del Boca et al., 2018) who investigate the role of grand-parenting in early child care in child outcomes. Our work complements these studies, but considers a longer time horizon and examines the effect of being bilingual on children's language, emotional, and pro-social development at various stages between the ages of three and fourteen years.

3 Data: The UK MCS

The data used in this analysis come from the UK MCS of approximately 19,000 children born between 2000 and 2001 (Plewis et al., 2007). Earlier UK birth cohort studies sampled babies born within a given week; on the other hand, the MCS has the advantage of capturing a birth cohort of infants who were born across a whole year. To date, six surveys have been conducted, at nine months, three, five, seven, eleven, and fourteen years old, resulting in a uniquely detailed portrait of children's development.

The MCS collects information about many diverse aspects of children's lives including: children's behavior, cognitive development, health, schooling, and living arrangements; parental employment and education; family income and poverty status; housing, neighborhood, and residential mobility; and social capital and ethnicity. The main unit of observation is the cohort member (the child), and the required information is collected from cohort members and the main respondent (typically the child's biological mother).⁵ For example, children's verbal reasoning test scores, which function as a measure of language development, are collected directly from children by trained interviewers, while questions about the children's socio-emotional and pro-social behavior are asked of parents.

⁵ The "main respondent" is defined as a person who answered the main interview questions regarding the cohort child. In the majority of cases, this was the biological mother.

3.1 Bilingualism in our estimation sample

Bilingualism is a multifaceted phenomenon with no commonly accepted definition. We adopt the perspective proposed by Kohnert (2010), who views bilinguals as “individuals who receive regular input in two or more languages during the most dynamic period of communication development - somewhere between birth and adolescence” (p. 457). We use information pertaining to the language spoken at home to construct a measure of bilingual status.⁶ The main respondent, usually the mother, is asked: “Is English the language spoken at home?” Children of respondents reporting “Yes – English only.” or “Yes – mostly English and sometimes other languages.” are classified as monolingual, whereas the children of respondents reporting “Yes – about half English and half other language.”, “No – mostly other, sometime English.”, or “No – other languages only.” are classified as being bilingual.⁷ The self-reported “language spoken at home” is not a perfect measure of bilingualism, since bilingualism is more than simply a language experience. Our data, however, do not include measures of English fluency of the parents, nor do we have precise measures of the amount of time for which each language is spoken.⁸ Consequently, we follow Locay et al. (2013) and capture “bilingualism” through the language spoken at home.

Information about the language status of the children constituting our sample is depicted in Figure 1. At age three, 56% of the children are exposed to a foreign language at home; however, by age 14, less than 30% of children continue to be bilingual.

Given our research interests, we focus on a sub-sample of UK born children, for each of whom at least one parent is foreign-born; these are the children most likely to be exposed to multiple languages at home.⁹ A total of 3,245 children (at age 9 months) meet these selection requirements. We further restrict the sample to cases with complete information on our language, emotional, and pro-social development measures and other control variables. These restrictions yield a sample of 1,993 children at age three; with subsequent attrition the sample was reduced to 1,682 at age five; 1,693 at age seven; 1,504 at age eleven; and 1,361 at age fourteen.¹⁰ In our robustness sub-section, we also report some additional checks performed

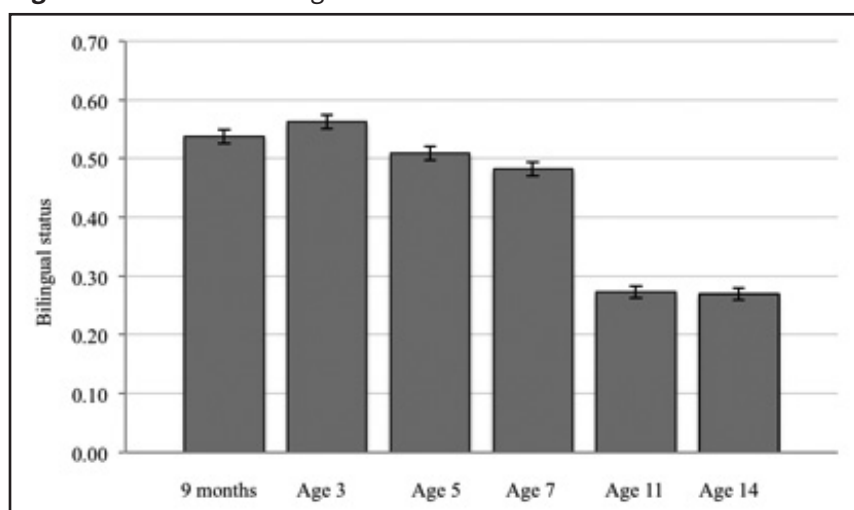
6 We assume that a bilingual child who attends full-time school, is given adequate exposure to both English and their other language. Research indicates that when children divide their learning time equally between the two languages, this results in 50% exposure to one language and 50% to the other; thus, the two languages develop like that of monolingual children (Genesee et al., 2004).

7 The survey questions on the language spoken at home in the first wave (child at nine months) are slightly different. In particular, the first wave provides only three options to the question: “Yes – English only”, “Yes – English and another language”, and “No – other languages only”. The first option is classified as monolingual and the second and third options are considered as bilingual. Our results are robust when compared to the alternative construction - the first two options classified as monolingual and the third as bilingual. We also based our definition on the findings from Ramrez (2016) that at 11 months of age, just before most babies begin to say their first words, babies from monolingual English households are specialized to process the sounds of English, and not the sounds of Spanish, an unfamiliar language. However, babies from bilingual Spanish-English households are specialized to process the sounds of both languages, Spanish and English.

8 Many different definitions of bilingualism have been used in the literature. See Lambert and Macnamara, 1969; Skutnabb-Kangas, 1981; Byram, 2004; and Fromkin et al., 2018. In particular, Skutnabb-Kangas (1981) also focuses on immigrant children, arguing that bilingualism has to do with person’s origin.

9 Since the entirety of the MCS sample children are born in the UK, all children in our sample are exposed to broadly the same institutional and cultural environment outside of the home. Moreover, restricting the sample to children with at least one foreign-born parent is helpful in reducing any heterogeneity associated with the fact that foreign-born parents may differ from native-born parents in ways that are non-random.

10 The longitudinal pattern of response in the MCS is complex, with attrition, re-entry, and a small number of late entrants who were eligible at wave 1 but were not included since they were not recorded on the register for Child Benefit. Child Benefit, a universal provision payable from the child’s date of birth, was used as the sampling frame for the MCS. Our regression estimates are weighted to account for non-response rates.

Figure 1 Children's bilingual status over time.

Source: The UK MCS. Authors' calculations are based on a bilingual sample.

across different sub-samples. In particular, we expect that having both parents come from the same country of birth exposes children to the parental language in a more consistent way than does having parents who both come from an English or non-English background.

We begin with an analysis by age group that exploits information for these children at every age in relation to which they are observed. We subsequently estimate dynamic models that rely on an unbalanced panel of 9,532 child-wave observations for those children aged three to fourteen for whom we have complete information on the key development measures and additional controls of interest.¹¹

3.2 Language and emotional development

The MCS provides a range of cognitive ability measures which are intended to be administered directly to children. We use age appropriate tests that come from the British Ability Scales (BAS). Our measure of language development comes from a series of tests of verbal reasoning and knowledge, and is constructed using the following assessments: the BAS naming vocabulary test taken at ages three and five; the BAS word reading test taken at age seven; the BAS verbal similarity test taken at age 11; and the BAS word activities test taken at age 14. The naming vocabulary score reflects expressive language skills, vocabulary (knowledge of nouns), ability to attach verbal labels to pictures, retrieval of names from long-term memory, and general level of language development (Hansen et al., 2012). The word reading score assesses English reading ability, while the verbal similarity score assesses verbal reasoning and verbal knowledge. Finally, the word activities assessment score measures the ability of children to understand the meaning of words. For each of these verbal reasoning tests, we use the age-standardized scores. To facilitate interpretation of the results, we transform the outcomes into z-scores. We are unable to consider cognitive outcomes on other cognitive tests (i.e., the school readiness and mathematical reasoning tests) since there are no consistent measures available over time.

¹¹ To maintain statistical power, in the analysis performed by child's age, we do not restrict our sample to children with non-missing information on all co-variates across all waves. Our unbalanced panel of 9,532 child-wave observations is further reduced to 6,310 observations due to the lag instrumental structure utilized in the GMM approach.

The emotional development of children is assessed using the Strength and Difficulties Questionnaire (SDQ) (Goodman, 2001), which is completed by both the main respondents and their partners. It comprises 25 questions encompassing five domains of behavior, namely: i) emotional issues: child complains of headaches, stomach aches/sickness, often seems worried, unhappy, nervous, or clingy in new situations; ii) conduct problems: child often has temper tantrums, fights with or bullies other children, and is often argumentative with adults; iii) hyperactivity: child is restless, over-active, cannot stay still for long, is constantly fidgeting, and is easily distracted; iv) peer problems: child tends to play alone, does not have at least one good friend, is not generally liked by other children, is picked on or bullied by other children, and gets on better with adults; and v) pro-social behavior: child is considerate of others' feelings, shares readily with others, is helpful if someone is hurt, upset, or ill, and is kind to younger children. Respondents indicate whether each item is "not true", "somewhat true", or "certainly true". Scores from the conduct problems, hyperactivity, emotional symptoms, and peer problems sub-scales are summed to construct a total SDQ score which varies between 0 and 40, with higher scores being indicative of more behavioral and emotional difficulties. Note that we exclude the pro-social behavior from the SDQ index since it is a positive trait and will be examined separately. The absence of pro-social behaviors is conceptually different from the presence of psychological difficulties (Goodman, 2001). Scores of the pro-social sub-scales range from 0 to 10 with a higher score indicating more engagement in pro-social behavior. The SDQ and pro-social scores are expressed as a normalized z-score.

3.3 Control variables

A major advantage of the MCS data is that they include detailed information about socio-economic characteristics, migration background, and parental inputs. Our socio-economic controls include the child's age, gender, ethnicity, birth weight, number of siblings, and the question of whether their biological mother lives in the household; also included are the mother's age and marital status, parental education levels, region of residence, and the family's poverty status (i.e., income being less than 60% of the median UK income).¹² Parental education at birth captures the additional time that educated parents, especially mothers, tend to spend with their child and also the fact that better educated parents are more likely to be fluent in English (Locay et al., 2013).

There is evidence that the language proficiency of one family member is positively associated with that of other family members, and that children's language proficiency is more highly correlated with maternal, rather than paternal, proficiency (Chiswick et al., 2005). Unfortunately, the MCS does not include measures of parental fluency in English. It does, however, provide information about migrant status, age at (UK) arrival, and country of origin, which may all be correlated with parental language proficiency and, consequently, with the children's propensity to speak English at home. In particular, we control for whether mothers and fathers come from a non-English-speaking country.¹³ We also control for whether migrant

¹² This is the official measure of poverty that is reported in the official Households Below Average Income Report (Department for Work and Pension, 2018), and is defined as living in a household with net equivalent income less than 60% of the UK median household income.

¹³ We used The World Almanac and Book of Facts (Park, 2005) to determine whether English was an official or predominant language in each country of origin. We classify the following as English-speaking countries: Australia, Canada, Barbados, Bermuda, Dominica, Ethiopia, Ghana, Guyana, Hong Kong, India, Ireland, U.S., Jamaica, Kenya, New Zealand, Nigeria, Saint Lucia, Saint Vincent and the Grenadines, Seychelles, Singapore, Trinidad and Tobago, and Zambia.

parents arrived in the UK before the age of 11 (as they are likely to have a higher degree of English proficiency).¹⁴ In our robustness section, we provide an additional sensitivity check, in which we exclude children whose parents arrived after 1997 (585 individuals arrived between 1997 and 2002), thus retaining a sample of children having parents who arrived between 1954 and 1996.

In addition, given that the survey provides information on the specific language used at home, such as Cantonese and Arabic, we experiment in constructing a measure indicating the language distance between the home language and English (see Isphording and Otten, 2014 and Isphording, 2015).¹⁵ However, given the dynamic nature of our research question, this time-invariant measure does not provide any useful insights and is not utilized in the current analysis.¹⁶

Finally, parents' investments in their children's development are captured in responses to detailed questions regarding the interactions that parents have with their children. Specifically, when children were nine months old, the main respondent was asked how important: i) talking; ii) cuddling; iii) stimulating; and iv) establishing regular sleeping and eating times were for the development of the child. The responses to these questions provide us with an insight into parents' approach to child rearing in infancy, and we control for this in our cross-sectional analysis at age three. We measure parental investments at older ages using information about how often parents read to their children or take them to the library.¹⁷

3.4 The bilingual gap in children's language, emotional and pro-social development

Table 1 reports the summary statistics of our unstandardized measures of development outcomes, along with other characteristics.¹⁸ There are pronounced differences between the two groups of children. Within their age cohort, bilingual children score much lower than their monolingual counterparts in the verbal reasoning exercise at both age three and five, but by age

¹⁴ The parents in our sample arrived between 1954 and 2002. Parents from non-English-speaking countries who arrive at an early age typically have English language skills comparable to migrant parents from English-speaking countries (Bleakley and Chin, 2008). Age at arrival may also have an impact on cultural assimilation. Those parents who arrived at older ages may differ in their values, views about parenting, etc., all of which may have an impact on the language and emotional development of their children.

¹⁵ Isphording (2015) shows that linguistic barriers caused by language differences play a crucial role in the determination of the destination-country language proficiency of immigrants. The argument is that immigrants face very different costs of language acquisition associated with their linguistic origin. Immigrant children learn English easier and faster if their language of origin is linguistically closer to that of the host country (Isphording and Otten, 2014). Hence, the cost of acquiring the host country language depends on the distance of the migrants' mother tongue from the dominant majority language, in our case English (Dustmann et al., 2003).

¹⁶ Linguistic distance may also be utilized through an instrumental variable (IV) strategy along with other covariates, such as age at arrival interacted with non-English-speaking background (see Bleakley and Chin, 2004). However, the exclusion restriction required to justify IV estimation is unlikely to hold good in our setting. Younger migrants are likely to differ from older migrants along non-language dimensions that also affect children's outcomes. Instead, we experiment by controlling for linguistic distance from both parents' language and English, rather than using indicators for having parents from non-English backgrounds. These results, available on request, overall reveal a similar pattern by child's age. Only three years olds whose parents speak a foreign language at home have a significantly higher probability of reporting emotional issues. They also tend to be less pro-social.

¹⁷ The level and quality of parental investments are usually proxied by the Home Observation Measurement of the Environment (HOME) scores that have been shown to be significantly correlated with later cognitive, health, and non-cognitive development (Todd and Wolpin, 2007; Cunha and Heckman, 2008). Similar constructs are available in our data but they are inconsistently measured over waves and therefore are not used in our analysis.

¹⁸ Corresponding information for native-born children is available upon request. Note that the scaling of the unstandardized verbal measure changes across waves, thereby making it not directly comparable across time.

Table 1 Summary statistics of selected variables

	Monolingual		Bilingual		Difference
Outcomes					
<i>Language development (verbal reasoning skills)</i>					
Age 3	50.42	[10.68]	38.87	[11.30]	-11.55***
Age 5	55.1	[10.69]	43.95	[11.65]	-11.15***
Age 7	115.47	[18.07]	114.88	[17.35]	-0.59
Age 11	60.15	[9.70]	56.98	[10.60]	-3.17***
Age 14	6.98	[3.38]	6.1	[2.80]	-0.88***
<i>Emotional development (SDQ score)</i>					
Age 3	9.11	[5.00]	11.18	[5.88]	2.07***
Age 5	6.89	[4.60]	8.49	[5.29]	1.61***
Age 7	6.87	[5.02]	8.26	[5.20]	1.39***
Age 11	7.07	[5.37]	7.71	[5.25]	0.64**
Age 14	13.3	[3.89]	14.35	[4.33]	1.05***
<i>Pro-social development</i>					
Age 3	7.40	[1.87]	7.46	[1.99]	0.06
Age 5	8.42	[1.66]	8.27	[1.80]	-0.15**
Age 7	8.63	[1.60]	8.46	[1.74]	-0.17**
Age 11	8.77	[5.37]	8.69	[1.63]	-0.08
Age 14	8.32	[3.89]	8.25	[1.95]	-0.08
Time-varying controls					
<i>Child's age in months</i>					
Age 3	37.62	[2.34]	38.24	[3.21]	0.62***
Age 5	62.67	[2.84]	62.63	[2.83]	-0.04
Age 7	86.96	[2.84]	87.12	[2.85]	0.16
Age 11	128.06	[5.68]	127.97	[5.74]	-0.09
Age 14	165.06	[5.42]	165.11	[5.45]	0.05
<i>Below 60% of the median UK income</i>					
Age 3	0.29	[0.45]	0.54	[0.50]	0.25***
Age 5	0.27	[0.45]	0.57	[0.50]	0.30***
Age 7	0.26	[0.44]	0.55	[0.50]	0.29***
Age 11	0.28	[0.45]	0.56	[0.50]	0.28***
9 months: Gross motor delay	0.11	[0.32]	0.14	[0.35]	0.03*
9 months: Fine motor delay	0.12	[0.33]	0.14	[0.35]	0.02
9 months: Communication gesture delay	0.41	[0.49]	0.43	[0.50]	0.03
<i>Parental inputs</i>					
9 months: Importance of stimulating the baby	0.36	[0.57]	0.58	[0.67]	0.22***
9 months: Importance of talking to the baby	0.15	[0.38]	0.31	[0.53]	0.16***
9 months: Importance of cuddling the baby	0.12	[0.34]	0.24	[0.51]	0.13***
9 months: Importance of regular sleep/feeding time	0.59	[0.75]	0.75	[0.82]	0.16***
Age 3: Frequency of the child being taken to the library	0.56	[0.78]	0.41	[0.73]	-0.15***
Age 3: Frequency reading to the child	3.31	[0.98]	2.67	[1.34]	-0.63***
Age 5: Frequency of the child being taken to the library	0.73	[0.67]	0.68	[0.72]	-0.05*
Age 5: Frequency reading to the child	2.45	[0.63]	2.31	[0.75]	-0.14***
Age 7: Frequency of the child being taken to the library	0.75	[0.66]	0.8	[0.72]	0.05
Age 7: Frequency reading to the child	2.11	[1.59]	2.36	[1.45]	0.26***

(Continued)

Table 1 Continued.

	Monolingual		Bilingual		Difference
Time invariant controls					
Male	0.52	[0.50]	0.5	[0.50]	-0.02
Birth weight	1.23	[0.75]	1.71	[1.05]	0.48***
Young migrant	0.77	[0.42]	0.52	[0.50]	-0.24***
Mother born non-English country	0.26	[0.44]	0.88	[0.33]	0.62***
Father born non-English country	0.27	[0.44]	0.88	[0.33]	0.61***
Child is white	0.72	[0.45]	0.12	[0.33]	-0.60***
Mother with higher qualification	0.43	[0.50]	0.32	[0.47]	-0.11***
Father with higher qualification	0.59	[0.49]	0.57	[0.50]	-0.01

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

seven the gap in language development between the two groups becomes insignificant. Bilingual children also score higher on the SDQ total difficulties score and lower on the pro-social score, but these differences tend to decline with age. Children whose parents speak a foreign language at home come mainly from non-white, ethnic, and lower educated households.

Parental investment at nine months, as measured by mother's attitudes toward child rearing, is significantly different between bilingual and monolingual mothers. Mothers of bilingual infants are more likely to believe that talking to, cuddling, and stimulating their babies and ensuring regular sleeping and feeding habits for infants are important. When bilingual children are aged three to five, mothers score lower on our parental activity measures, whereas, in contrast, when bilingual children are aged seven, parents tend to spend more time reading to their child and visiting the library. These results suggest that there is some variation in parental practices over time.

4 Methods

Our primary objective is to understand whether – relative to their monolingual peers – there is a gap in the language, emotional, and pro-social development of children exposed to a foreign language at home, and how this gap evolves over time. We address this issue by estimating child development production functions using the approach developed by Todd and Wolpin (2003, 2007) and applied in Fiorini and Keane (2014), Del Boca et al. (2017), and Del Bono et al. (2016). Children's development is assumed to be a cumulative process that depends on both contemporaneous and historical family investments, as well as children's skill endowments. Our focus is on the developmental consequences of families' decisions to raise their children in a bilingual home environment.

There are well-known econometric challenges involved in generating unbiased estimates of production function parameters along with observational data. The first hurdle arises because investments in children are not exogenous, but instead result from the active choices that parents make when trying to maximize their children's human development given the constraints they face. This would not necessarily be a problem if comprehensive data on all

relevant inputs (e.g., parental, school, and community investments; child endowments; etc.) pertaining to child development were to be observed (see Todd and Wolpin, 2003); however, this is rarely, if ever, the case. These unobserved inputs (e.g., children's innate ability) are almost certainly correlated with certain other inputs (e.g., reading to children) which are observed by researchers. This correlation results in the usual omitted variable bias problem. The second problem stems from the threat that both structural (e.g., simultaneity and mutual causality) and statistical endogeneity (e.g., unobserved heterogeneity and measurement error) pose to causal estimation of the effect of parental investments on children's development.

Our empirical strategy is to account for the cumulative nature of child development using value-added specifications which control for measures of lagged development (see Hanushek et al., 2009 and Harris, 2010 for details). Specifically, our model of children's development is given by:

$$C_{iw} = \beta B_{i,w-1} + \lambda C_{i,w-1} + \mathbf{X}'_{iw} \gamma + \varepsilon_{iw} \quad (1)$$

where C_{iw} denotes the language, emotional, and pro-social development for child i in waves 2, 3, 4, 5, and 6 (or at child's age 3, 5, 7, 11, and 14), $B_{i,w-1}$ is an indicator of bilingual status in wave ($w-1$), and \mathbf{X}'_{iw} is a vector of child- and family-specific covariates. The advantage of using lagged values (instead of contemporaneous values) of bilingual status is that we can ensure that the lagged values precede the child's performance on the verbal language test and development of emotional difficulties, thereby ruling out reverse-causality concerns. We include a rich list of factors which contribute to child development such as child's age, gender, ethnicity, number of siblings and birth weight, maternal age, parental education measured at nine months, the presence of the biological mother in the household, parents' non-English background, young migrant indicator, reading and library visitation frequency, poverty, and regional indicators. Finally, $C_{i,w-1}$ captures lagged development, while λ is a persistence (autoregressive) parameter that links development across periods, such that $|\lambda| < 1$. Including $C_{i,w-1}$ in the model not only captures persistence in language and emotional development, but also controls for unobserved ability (see Todd and Wolpin, 2003). Under certain conditions, the lagged development measure effectively acts as a sufficient statistic for all historical time-varying inputs (see Sass et al., 2014).

Ordinary Least Square (OLS) estimation of Eq. (1) produces unbiased estimates of the child development production function, for as long as ε_{iw} is i.i.d. There may, however, also be unobserved differences in the rate at which children acquire language skills or develop emotionally. This implies that the error term is best modeled as $\varepsilon_{iw} \equiv \mu_i + v_{iw}$, where μ_i captures child-specific differences in the rate of development. Even if we make the strong assumption that there is no correlation between family inputs and children's unobserved ability, OLS estimation of Eq. (1) will not generally result in unbiased estimates of the child development production function, because lagged development $C_{i,w-1}$ will depend on μ_i , causing it to be correlated with ε_{iw} . Unfortunately, fixed-effect estimation does not resolve the issue (Nickell, 1981).

Consequently, in addition to using OLS, we also undertake instrumental variable estimation of Eq. (1) using a generalized method of moments (GMM) framework. This results in consistent estimates even in the presence of child-specific unobserved heterogeneity as further lags of development measures which use two or more period lags of dependent variable and/or of X -variable, and which are valid instruments for $C_{i,w-1}$ (see Arellano and Bond, 1991 and Blundell and Bond, 1998). This estimation strategy has been successfully applied in related papers which are focused on skill formation (Del Bono et al., 2016) and the education

production function (Andrabi et al., 2011). Specifically, we estimate Eq. (1) using a system of equations – referred to as “system-differences and levels” GMM – which includes Eq. (1) along with an additional equation that purges the child-specific effect μ_i from the model:

$$\Delta C_{iw} = \beta \Delta B_{i,w-1} + \lambda \Delta C_{i,w-1} + \Delta \mathbf{X}'_{iw} \gamma + \Delta v_{iw} \quad (2)$$

where Δ prefix denotes the change from one period to the next, i.e., $\Delta C_{iw} = C_{iw} - C_{i,w-1}$, $\Delta B_{iw} = B_{iw} - B_{i,w-1}$, $\Delta \mathbf{X}_{iw} = \mathbf{X}_{iw} - \mathbf{X}_{i,w-1}$, $\Delta C_{i,w-1} = C_{i,w-1} - C_{i,w-2}$, and $\Delta v_{iw} = v_{iw} - v_{i,w-1}$.

The instruments used in this system are the two-period lagged measures of language and emotional difficulties (and pro-social behavior), the two-period lag of bilingual status along with contemporaneous inputs (child’s age in months and birth weight) for the difference as per Eq. (2), and past inputs (presence of the natural mother in the household) along with invariant inputs such as child’s gender, ethnicity, and mother’s and father’s education at child’s birth for the levels as per Eq. (1).¹⁹ Note that the variability exploited by the system-GMM estimation comes from those children who switch their bilingual status in the observation period.²⁰ Hence, we need a sufficient number of children to change their bilingual status between survey waves in order to accurately estimate the effects. Although such variation may not seem very intuitive, it is established that transition to primary school is a key moment, when linguistic habits are subject to change; the combination of a stronger dominant language and social differences may lead many children to reject their home language (see Hundt, 2006). In our data, we observe 700 children who switched from bilingual to monolingual status, while 235 children switched from monolingual to bilingual status over the observed period.²¹ The dynamic GMM estimator is unbiased and consistent under the assumption that there is no second-order serial correlation in the error term. To test this assumption, we use a test procedure outlined in Arellano and Bond (1991). Specifically, results from two tests of the validity of the instruments (the Sargan–Hansen test of over-identifying restriction) and autocorrelation (the AR(2) process) are performed.²²

5 Results and Discussion

We begin our empirical analysis with age-specific models which provide descriptive evidence of evolution in the bilingual development gap from an early childhood (age of three) to

19 When experimenting with different instruments, we look for a proliferation of instruments that may overfit the endogenous variables, and we ensure that the model passes both the test for instrument validity (i.e., the Sargan–Hansen test) and the test for second-order serial correlation (see Arellano and Bond, 1991).

20 The GMM can be used for periods of $T \geq 3$ since identification comes from large N , and not from large T . Arellano and Bover (1995) and Blundell and Bond (1998) demonstrate that the system-GMM approach – by adding additional moment restrictions – permits lagged first differences to be used as instruments in the levels equations, and this corrects for any bias that would emerge as a consequence of using the standard GMM estimator.

21 In Table A1 in Appendix, we report the characteristics of those children who switch bilingual status. Children who switch exclusively to the English language, as early as the age of five, score relatively higher in the verbal reasoning exercise, when compared to those who always stayed bilingual or moved to bilingual at age of seven. A feature of our specification is that there is no differentiation in the effect of bilingualism among those who move into versus those who move out of bilingual status.

22 One limitation of our data is the irregularly spaced survey design; this may have an impact on the validity of the Arellano–Bond tests. Our waves refer to a longer period and there may be some unobservable factors that we have not fully captured. To address the missing data issue, Millimet and McDonough (2017) incorporate imputation of the missing covariates. As an alternative, one might consider imputation of the missing lagged dependent variable. However, this introduces additional complications in that the measurement error, representing the deviation between the actual lagged dependent variable and the imputed variable, is likely to depend on the deviation between the covariates from period $t-1$ and the covariates used to impute the missing value (see Millimet and McDonough, 2017). We experimented with the Millimet and McDonough (2017) estimator; however, due to the smaller sample size the estimated effects are less precise.

adolescence (age of 14). All estimates are weighted to account for the non-response rate and standard errors are clustered at the family level. We then turn to our dynamic OLS and GMM estimates of the value-added child development production function by pooling all data waves together. In Subsection 5.3, we examine whether there is evidence that the production functions of language, emotional, and pro-social skills are different for subgroups of the population. Subsection 5.4 documents whether there are corresponding patterns of disparities in maternal behavior and home environments, and finally, in Subsection 5.5 we test the robustness of our estimates in relation to several sample restrictions.

5.1 The bilingual gap in children's development by age

We first investigate how the bilingual gap evolves from early childhood to adolescence by estimating Eq. (1) separately by age group. These results provide evidence of the way in which the bilingual gap in development evolves as children mature. We report the estimated bilingual gap separately by age, for language, emotional development, and pro-social behavior as measured by the standardized verbal reasoning test scores (Table 2), the SDQ score (Table 3), and the pro-social score (Table 4). We consider three specifications: Panel (A) in Tables 2–4 includes only basic controls (child's age in months, mother's age at birth, and gender and ethnicity); Panel (B) in Tables 2–4 includes the full set of controls, but omits measures of lagged language and emotional development, or pro-social behavior; and Panel (C) in Tables 2–4 reports the results of our cumulative value-added specification that includes lagged development measures along with our full set of controls.²³ Comparing results with and without measures of lagged development helps us to gauge the importance of using the value-added specification to account for previous, unobserved family inputs into child development. Children are interviewed approximately every three years; therefore, the coefficients on lagged bilingual status and other family inputs can be interpreted as an effect arising from application of the input over a three-year interval.

The estimates reported in Table 2, clearly point to a strong negative correlation between language spoken at home during early childhood and subsequent language development. We find that three-year-old children whose parents speak a foreign language at home have significantly lower verbal reasoning test scores than their monolingual peers. Other things being equal in a very basic specification (Panel (A) in Tables 2–4), speaking a foreign language at home when the child was nine months old reduced the child's language development, measured at age three, by 0.63 standard deviations. Adding a comprehensive set of controls reduces the estimated bilingual deficit in language development relative to that of an otherwise similar monolingual child to 0.35 standard deviations (see Panels (B) and (C) in Tables 2–4). In comparison, the negative impact of bilingualism on language is greater than the effect of low socio-economic background – a nine-month-old child who is living in poverty can be expected

²³ The full set of controls included in Panels (B) and (C) at each age are slightly different depending on data availability but tend to include child's age in months, gender, ethnicity and birth weight, mother's age and marital status, mother's and father's highest qualification measured when the child is nine months old, an indicator for young migrants, non-English speaking background, whether the biological mother present in the household, number of siblings, frequency of reading to the child or library visits, poverty indicator, regional controls, along with time-varying characteristics measured at period $a - 1$, such as poverty indicator, number of siblings, marital status, and presence of biological mother.

Table 2 The effects of bilingualism on language development; the selected coefficient is child’s age

z-Verbal reasoning skills					
(I) 2–6	Age 3	Age 5	Age 7	Age 11	Age 14
<i>Panel (A): OLS with basic controls</i>					
Bilingual _{w-1}	-0.629*** (0.053)	-0.587*** (0.057)	0.069 (0.059)	-0.246*** (0.064)	-0.153** (0.059)
R ²	0.276	0.187	0.022	0.050	0.033
<i>Panel (B): OLS basic + full controls</i>					
Bilingual _{w-1}	-0.352*** (0.062)	-0.379*** (0.067)	0.096 (0.071)	-0.037 (0.071)	-0.077 (0.062)
R ²	0.379	0.308	0.111	0.176	0.092
<i>Panel (C): OLS basic + full controls + past endowments</i>					
Bilingual _{w-1}	-0.345*** (0.061)	-0.199*** (0.062)	0.219*** (0.067)	-0.022 (0.068)	-0.085 (0.062)
z-Verbal reasoning skills _{w-1}		0.403*** (0.024)	0.294*** (0.025)	0.255*** (0.030)	0.202*** (0.026)
z-SDQ score _{w-1}		-0.039* (0.021)	-0.182*** (0.026)	-0.030 (0.028)	-0.043 (0.028)
R ²	0.394	0.418	0.217	0.233	0.135
N	1993	1682	1693	1504	1361

Notes: Standard errors, clustered at the family level, are in parentheses. All regressions are weighted to account for the non-response rate. Each panel corresponds to a separate specification, where the bilingual effect with basic controls (child’s age in months, gender and ethnicity, and mother’s age) is reported in Panel (A); Panel (B) additionally includes child’s birth weight, mother’s age and marital status, mother’s and father’s highest qualification obtained when the child is nine months old, a young migrant indicator, mother and father coming from non-English speaking background, biological mother present in the household, child’s number of siblings, frequency of reading to the child or library visits, poverty indicator, regional controls along with time-varying characteristics measured at period *w* -1, such as a poverty indicator, child’s number of siblings, mother’s marital status, and presence of biological mother; however, it omits lagged endowment effects; Panel (C) in addition to the full set of controls includes lagged input measures for language and emotional development.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

to have a verbal reasoning score that is 0.14 standard deviations below that of a non-poor child at age three.²⁴

The penalty for not speaking English exclusively at home is reduced at age five, once we fully control for the observable characteristics and past developmental outcomes; however, bilingual children continue to score lower on average in comparison to their monolingual

24 The full set of estimated coefficients are available upon request. These estimates additionally reveal that the education of the mother and father, measured at the time of the child’s birth, has a significant and positive association with language development at age three. For example, having a mother with a higher education degree is associated with a 0.19 standard deviation increase in the verbal reasoning score, as opposed to having a mother with a lower qualification. Children with gross, fine motor development and communication delays measured at nine months have not caught up developmentally by age three. The effect of having parents who arrived in the UK at younger ages results in a significant positive increase in verbal development at age three by around 0.13 standard deviations. Parental inputs, as measured by the frequency with which the mother or father reads to the child and/or visits the library, significantly impacts the verbal outcomes of five-year-old children. Overall, adding a comprehensive set of controls reduces the estimated coefficient on current bilingual status, and in some specifications, the effect is rendered statistically insignificant at the conventional levels.

Table 3 The effects of bilingualism on emotional development; the selected coefficient is child's age

(I) 2–6	z-SDQ score				
	Age 3	Age 5	Age 7	Age 11	Age 14
<i>Panel (A): OLS with basic controls</i>					
Bilingual _{w-1}	0.219*** (0.058)	0.055 (0.057)	0.091 (0.056)	0.049 (0.065)	0.081 (0.072)
R ²	0.086	0.067	0.055	0.023	0.022
<i>Panel (B): OLS basic + full controls</i>					
Bilingual _{w-1}	0.109 (0.069)	-0.049 (0.073)	-0.031 (0.066)	-0.056 (0.072)	-0.038 (0.078)
R ²	0.165	0.126	0.118	0.083	0.084
<i>Panel (C): OLS basic + full controls + past endowments</i>					
Bilingual _{w-1}	0.105 (0.068)	-0.106* (0.061)	-0.043 (0.054)	-0.095 (0.060)	0.006 (0.071)
z-Verbal reasoning skills _{w-1}		-0.033 (0.023)	-0.047** (0.022)	-0.110*** (0.024)	-0.016 (0.032)
z-SDQ score _{w-1}		0.493*** (0.025)	0.561*** (0.023)	0.520*** (0.025)	0.494*** (0.031)
R ²	0.175	0.354	0.420	0.386	0.276
N	1993	1682	1693	1504	1361

Notes: Standard errors, clustered at the family level, are in parentheses. All regressions are weighted to account for the non-response rate. Each panel corresponds to a separate specification where the bilingual effect with basic controls (child's age in months, gender and ethnicity, and mother's age) is reported in Panel (A); Panel (B) additionally includes child's birth weight, mother's age and marital status, mother's and father's highest qualification obtained when the child is nine months old, a young migrant indicator, mother and father coming from non-English speaking background, biological mother present in the household, child's number of siblings, frequency of reading to the child or library visits, poverty indicator, regional controls along with time-varying characteristics measured at period $w-1$, such as a poverty indicator, child's number of siblings, mother's marital status, and presence of biological mother; however, it omits lagged endowment effects; Panel (C) in addition to the full set of controls includes lagged input measures for language and emotional development.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

counterparts even if we control for past development levels. The gap is statistically significant at the 1% level. Other things being equal, the results suggest that, on average, bilingual children's verbal reasoning scores are 0.20 standard deviations below that of monolingual children. This provides suggestive evidence that five-year-old bilingual children continue to fall behind in their expressive language and knowledge of names. Consistent with the results of the literature available from Cunha and Heckman (2008) and Dickerson and Popli (2016), past language endowment has a positive and significant effect on language development as measured at a later age. For example, an increase of one standard deviation in the verbal reasoning test score at age three is associated with greater language development (0.40 of a standard deviation) at age five.

Language development in bilingual children, at the age of seven, in the basic specification, is found to be statistically equivalent to that in monolingual children in specifications that omit the prior developmental outcomes (see Panels (A) and (B) in Table 2). However, once we control

Table 4 The effects of bilingualism on pro-social development; the selected coefficient is child's age

(l) 2–6	z-Pro-social score				
	Age 3	Age 5	Age 7	Age 11	Age 14
<i>Panel (A): OLS with basic controls</i>					
Bilingual $w-1$	-0.114** (0.058)	0.051 (0.062)	-0.014 (0.060)	-0.070 (0.070)	-0.048 (0.070)
R^2	0.020	0.034	0.021	0.023	0.028
<i>Panel (B): OLS basic + full controls</i>					
Bilingual $w-1$	-0.089 (0.072)	0.065 (0.086)	0.035 (0.070)	0.014 (0.081)	-0.030 (0.073)
R^2	0.053	0.055	0.049	0.046	0.048
<i>Panel (C): OLS basic + full controls + past endowments</i>					
Bilingual $w-1$	-0.088 (0.071)	-0.002 (0.078)	0.039 (0.063)	0.071 (0.075)	0.004 (0.066)
z-Pro-social score $w-1$		0.306*** (0.026)	0.399*** (0.026)	0.395*** (0.032)	0.414*** (0.030)
R^2	0.062	0.148	0.225	0.183	0.231
N	1993	1682	1693	1504	1361

Notes: Standard errors, clustered at the family level, are in parentheses. All regressions are weighted to account for the non-response rate. Each panel corresponds to a separate specification where the bilingual effect with basic controls (child's age in months, gender and ethnicity, and mother's age) is reported in Panel (A); Panel (B) additionally includes child's birth weight, mother's age and marital status, mother's and father's highest qualification obtained when the child is nine months old, a young migrant indicator, mother and father coming from non-English speaking background, biological mother present in the household, child's number of siblings, frequency of reading to the child or library visits, poverty indicator, regional controls along with time-varying characteristics measured at period $w-1$, such as a poverty indicator, child's number of siblings, mother's marital status, and presence of biological mother; however, it omits lagged endowment effects; Panel (C) in addition to the full set of controls includes lagged input measures for pro-social development.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

for the measures of prior language and emotional development, we find that seven year olds who were bilingual at age five have significantly higher language development (0.22 standard deviations) than monolingual children, whether or not they continue to be bilingual themselves at age seven.²⁵ To obtain an understanding of how important these estimates are, we can compare them to the effect of parental education. At the age of seven, having a mother and father with a higher education degree is associated respectively with a 0.10 and 0.11 standard deviation advantage in language development, relative to having parents with lower qualifications. The gap associated with being bilingual (measured via the one-period lag of bilingualism) is nearly twice as large. The persistence in language outcomes, as measured by past verbal reasoning skills, is also statistically significant and supports the notation of self-productivity – skills acquired during previous periods enhance skill formation at a later period (Cunha et al., 2006).

25 Recall that in the verbal reasoning test taken at age seven (the BAS-Word reading test), the child must correctly pronounce words within locally accepted standards, with emphasis on the correct syllable or syllables.

These results reported at the age of seven are consistent with Mumtaz and Humphreys (2001), who show that bilingual children are better at reading as compared to their monolingual counterparts. This suggests a possible “transfer” of first language literacy skills to the development of reading in a second language. It also supports the view that bilingual reading development may aid the acquisition of certain literacy skills such as phonological awareness and memory, and the regular word-reading ability that accumulates over time.²⁶

At age 11, the advantage in the verbal reasoning tests for children who were bilingual at the age of seven (measured via the one-period lag of bilingualism) is not statistically significant once we fully control for child characteristics, parental characteristics, and children’s past developmental outcomes. These results are mirrored for bilingual children at age 14 with a decrease in verbal reasoning scores of 0.09 of a standard deviation, and the effect is insignificant. There is a statistically significant persistence in language development over time. Interestingly, the persistence in language development diminishes as children age; also, the negative association between having emotional difficulties and language development – as measured by the SDQ scores – which is observed at early ages is no longer evident once children reach adolescence.

Our results for emotional development, as measured by the standardized SDQ, are presented in Table 3. We find that the discrepancies in SDQ scores between bilingual and monolingual children, originally explained in Table 1, are insignificant once we fully control for observable characteristics and past endowments (see Panels (B) and (C) in Table 1). In other words, being bilingual has no significant impact on the emotional development of children, regardless of their age. Children’s past language development is found to be associated with their emotional development, however. An increase of one standard deviation in the past language score is associated with a reduction of 0.11 of a standard deviation in emotional difficulties at age eleven (see Panel (C) in Table 3). There is a strong persistence in emotional development over time; lagged SDQ scores are highly predictive of current SDQ scores. For example, a one standard deviation increase in emotional difficulties at age five is associated with a 0.56 standard deviation increase in emotional difficulties at age seven.

We observe similar results for pro-social development (see Table 4). The pro-social development of bilingual children is statistically equivalent to monolingual children in the specification that controls for past endowments of pro-sociality (see Panel (C) in Table 4). The persistence in pro-social behavior is significant and stable over time. Specifically, other things being equal, a one standard deviation increase in the lagged pro-social score increases current pro-social behavior by 0.30–0.40 standard deviations.

In summary, results in Tables 2–4 reveal several important findings. First, there is a negative relationship between bilingualism and early language development. The magnitude of this early disadvantage is significant, and consistent with some previous studies. In a more general context, Lang and Sepulveda (2007) find a black–white gap in cognitive tests, ranging from 0.20 to 0.50 of a standard deviation after controlling for a myriad of mother and family controls. The start of schooling, however, appears to attenuate the language gap and by age 14 the difference in early language development is insignificant. We find that there are no

²⁶ Mumtaz and Humphreys (2001) investigate the impact of Urdu as a first language on learning to read in English as a second language. All students were tested individually over a period of eight weeks. Bilingual children were also found to have an advantage in phonological awareness at the earliest stages of reading, as compared to monolingual children.

significant differences in children's emotional development and pro-social behavior. Finally, our estimates emphasize that prior language and socio-emotional inputs have a noticeable influence on child development.

5.2 Dynamic estimates of the bilingual gap in children's development

Results from the cumulative value-added models which were estimated via OLS and dynamic GMM on our unbalanced panel of 9,532 child-wave observations are presented in Table 5. Columns (1) and (2) in Table 5 report the OLS estimates of Eq. (1), without and with past language and emotional endowments, respectively. Column (3) in Table 5 shows the system GMM estimator that combines the levels and the first difference equation.²⁷

Omitting the previous language and emotional development and pooling all ages together (Column (1)), it is discovered that the lagged bilingual status is associated with a decrease in children's verbal language scores (0.10 of a standard deviation), other things being equal.²⁸ When past development measures are controlled, the estimated coefficient is smaller and not statistically different from zero (Column (2)). There is persistence in language development that is statistically significant; a one standard deviation increase in the lagged verbal reasoning scores is associated with an increase in the current verbal score of 0.28 of a standard deviation. In general, as in our age-specific analysis, the coefficient on past language skills is well below one, indicating that persistence in language achievement – while significant – is far from perfect. Lower persistence in the measures of language development indicates that learning is lost at a greater rate over time. Past emotional development index correlates negatively with language development.

The system GMM estimates in Column (3) allow for the possibility of detecting the presence of endogenous feedback effects, i.e., parents may adjust their behavior and resource decisions, as governed by feedback on children's outcomes. In the presence of such feedback effects, we expect to observe both a reduction in the persistence parameter, λ , and an increase in the bilingual coefficients (see Andrabi et al., 2011 and Del Bono et al., 2016). We do not find evidence which suggests that feedback effects may be present. The point estimates of the coefficient on past language skills increases from 0.28 in the OLS (Column (2)) to 0.60 in the GMM specification (Column (3)). The coefficient which is associated with past bilingual status increases in absolute value; however, it remains insignificantly different from zero. The system GMM standard errors are much higher than their OLS counterpart.²⁹ A natural question is how these dynamic estimates compare to the different age effects that are reported in Table 2. We keep in mind that the dynamic panel GMM estimator identifies the bilingual effect using children who switch between bilingual and monolingual status over time. The insignificant but positive coefficient associated with the past bilingual status suggests that there is no penalty for children who switch their bilingual status.

²⁷ The specifications also control for the differences in the survey time by including dummies for waves in regressions.

²⁸ In a specification that includes mother fixed-effects (not reported, but available upon request), the estimates are of a similar order of magnitude as those from the OLS model. This suggests that bilingual differences are not simply the product of unobserved variation in time-invariant family characteristics (see De Haan et al., 2014 for similar discussion.).

²⁹ In terms of precision, this estimator appears much more satisfactory than difference-GMM. This is because lagged differences are stronger instruments for lagged levels (whereas lagged levels are weak instruments for lagged differences).

Table 5 The effects of bilingualism on language, emotional, and pro-social development – OLS and GMM dynamic estimates

	(1)	(2)	(3)
	OLS no past endowments	OLS with past endowments	System GMM effects
<i>Panel (A): Outcome: z-Verbal Reasoning skills</i>			
Bilingual _{w-1}	-0.102*** (0.036)	-0.032 (0.031)	0.040 (0.106)
z-Verbal reasoning skills _{w-1}		0.278*** (0.014)	0.602*** (0.218)
z-SDQ score _{w-1}		-0.092*** (0.013)	-0.743** (0.352)
Year dummies	Yes	Yes	Yes
Hansen <i>p</i> -value			0.243
Arellano–Bond test AR(2) <i>p</i> -value			0.152
<i>Panel (B): Outcome: z-SDQ score</i>			
Bilingual _{w-1}	-0.006 (0.037)	-0.049* (0.027)	-0.018 (0.048)
z-Verbal reasoning skills _{w-1}		-0.066*** (0.011)	-0.069 (0.093)
z-SDQ score _{w-1}		0.490*** (0.014)	0.266* (0.155)
Year dummies	Yes	Yes	Yes
Hansen <i>p</i> -value			0.551
Arellano–Bond test AR(2) <i>p</i> -value			0.753
<i>Panel (C): Outcome: z-Pro-social score</i>			
Bilingual _{w-1}	-0.001 (0.025)	0.014 (0.021)	-0.028 (0.033)
z-Pro-social score _{w-1}		0.343*** (0.019)	0.287* (0.157)
Year dummies	Yes	Yes	Yes
Hansen <i>p</i> -value			0.344
Arellano–Bond test AR(2) <i>p</i> -value			0.238
<i>N</i> (Observations)	6310	6310	6310
<i>N</i> (Individuals)			2269

Notes: Robust standard errors, which are clustered at the family-level, are placed in parentheses. Columns (1) and (2) report OLS estimates of Eq. (1) without and with past language and emotional endowments, respectively. Column (3) reports the system GMM estimator which assumes that effects have constant correlation with the inputs. All estimates include the full set of controls and time dummies. Verbal reasoning skills refer to language development and the SDQ score refers to emotional development.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

Consistent with our age-specific results, using the dynamic GMM model, we find no effect of bilingualism on the emotional and pro-social development of children. The point estimate of z-SDQ score_{*a*-1} is 0.27 of a standard deviation in the system GMM specification, representing

a decrease with respect to the OLS results reported in Column (2). Similarly, the coefficient on past pro-social behavior decreases from 0.34 in the OLS specification (Column (2)) to 0.29 in the system GMM (Column (3)).

The consistency of the GMM estimates hinges upon the assumption that there is no serial correlation in the unobserved individual level, time-varying determinants. The P -values of the Sargan–Hansen test statistic of over-identifying restrictions indicate that we cannot reject the null hypothesis that the instruments as a group are valid. This constitutes evidence that the models are well specified. The subsequent P -values of the Arellano–Bond AR(2) test, reported in the bottom of Column (3) in Table 5, are in the range 0.15–0.75 for the three outcomes, showing little evidence for existence in serial correlation; hence, second order lags are likely to be appropriate instruments in the dynamic specifications.

Consistent with previous results, the dynamic specification also illustrates a significant persistence in language, emotional, and pro-social behavior over time. In particular, about 30% of pro-social development persists over time – a one standard deviation increase in the lagged pro-social score increases the current pro-social behavior by 0.29 of a standard deviation (Panel (C)) – while more than 0.60 of a standard deviation of verbal reasoning skills persists across ages. This finding, which is also emphasized by Del Bono et al. (2016) and Fiorini and Keane (2014), is in line with the idea that production functions for cognitive and non-cognitive skills are very different (Cunha et al., 2006).³⁰

5.3 Heterogeneity

Next, we explore whether there is any heterogeneity in the bilingual effects reported so far. First, we consider whether the effects of bilingualism on language, emotional, and pro-social development differ by the children's gender. This additional information is important for identifying whether, for example, gender-specific strategies might be more effective in reducing language difficulties that arise from the language spoken at home.³¹ The results, reported in Table 6, are based on age models with full set controls, including past endowments as reported in Panel (C) in Tables 2–4. We find that the gap in language development at an early age is not significantly different between boys and girls. Other things being equal, bilingualism, by age three, decreases verbal reasoning scores by 0.35 of a standard deviation for girls and by 0.32 of a standard deviation for boys. The differences become more pronounced as the children age. As indicated by the results in Table 6, boys mainly drive the bilingual language advantage at age seven; the lagged bilingual status is positive and significant at the 5% level. However, by age 14, boys are back to having lower language development. Boys' underachievement at that age, which corresponds to Key Stage 3, is in line with some official figures. For example, international comparisons of 15-year-

³⁰ We assess the sensitivity of our results by re-estimating our dynamic and child age-specific models using a balanced panel of children, with no missing information in the main family background variables over the five observed waves; that leaves us with 926 children, from whom 4,630 child-year observations could be gathered. We do not observe any substantive change in our bilingual coefficients. The results are available on request.

³¹ Previous research shows a gender gap in children's performance. For example, boys' attitudes towards reading and writing, the amount of time they spend in reading, and their achievement in literacy are all poorer than of the corresponding values for girls (All-Party Parliamentary Literacy Group Commission, 2012). The literature also suggests that gender differences do exist in the maintenance of home language. While girls are more likely to possess higher social skills and academic competence, boys often have more problem behaviors (Margetts, 2005). Furthermore, our estimates, not reported but available on request, consistently suggest that boys are more likely to experience emotional difficulties and that their verbal language development is well below the language development of girls.

Table 6 Heterogeneous effects of bilingualism, by children's gender

	Age 3	Age 5	Age 7	Age 11	Age 14
<i>Panel (A): Outcome: z-Verbal Reasoning skills</i>					
Boys					
Bilingual _{w-1}	-0.320*** (0.085)	-0.207** (0.090)	0.242** (0.110)	0.090 (0.101)	-0.200** (0.094)
<i>N</i>	1012	840	832	731	662
<i>R</i> ²	0.381	0.425	0.227	0.223	0.157
Girls					
Bilingual _{w-1}	-0.352*** (0.087)	-0.179** (0.084)	0.148* (0.085)	-0.155* (0.089)	0.007 (0.083)
<i>N</i>	981	842	861.000	773	699
<i>R</i> ²	0.424	0.440	0.224	0.268	0.132
<i>Panel (B): Outcome: z-SDQ score</i>					
Boys					
Bilingual _{w-1}	0.145 (0.096)	-0.165* (0.092)	-0.060 (0.080)	-0.187** (0.088)	-0.076 (0.105)
<i>N</i>	1012	840	832	731	662
<i>R</i> ²	0.175	0.357	0.442	0.427	0.323
Girls					
Bilingual _{w-1}	0.093 (0.100)	-0.041 (0.080)	-0.005 (0.074)	-0.009 (0.082)	0.124 (0.097)
<i>N</i>	981	842	861	773	699
<i>R</i> ²	0.201	0.356	0.409	0.358	0.261
<i>Panel (C): Outcome: z-Pro-social score</i>					
Boys					
Bilingual _{w-1}	-0.126 (0.102)	0.092 (0.122)	0.006 (0.099)	-0.001 (0.108)	0.069 (0.098)
<i>N</i>	1012	840	832	731	662
<i>R</i> ²	0.076	0.136	0.184	0.187	0.256
Girls					
Bilingual _{w-1}	-0.021 (0.102)	-0.093 (0.101)	0.076 (0.078)	0.184* (0.103)	0.099 (0.083)
<i>N</i>	981	842	861	773	699
<i>R</i> ²	0.067	0.153	0.256	0.192	0.217

Notes: Standard errors, which are clustered at the family level, are in parentheses. The specifications include the same controls as in Panel (C) in Tables 2–4. Verbal reasoning skills refer to language development and the SDQ score refers to emotional development.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

olds' Programme for International Student Assessment (PISA) results show that girls do better in reading than boys across all of the OECD countries (Bradshaw et al., 2009).

In the case of emotional development, it is mainly boys that are found to benefit from their bilingual background, though the effect varies by age. Specifically, at ages five and eleven,

the advantage in emotional development for boys is statistically significant when compared to girls. Other things being equal, speaking a foreign language at home when the child was seven years old reduces the total difficulties score by 0.19 of a standard deviation. No significant gap in the emotional development of girls is evident. To gain a better understanding of the domains underlying the emotional development result, we focus on the role of bilingualism by looking at the five SDQ domains separately, rather than at the single SDQ score. Results, available upon request, indicate a significant difference in conduct problems and hyperactivity for boys. Specifically, lagged bilingual status reduces conduct problems and hyperactivity outcomes by 0.18 and 0.26 of a standard deviation, respectively. Finally, for both boys and girls, speaking a foreign language at home does not associate with significantly different pro-social behavior. Although these results should be interpreted with caution due to the low sample sizes, explaining why the emotional development of boys benefits from bilingual status deserves more investigation. It could be hypothesized that the environmental and cultural background may influence children's social competence differently among boys and girls. We leave gender differences in emotional development between bilingual boys and girls to a potential future research.

Furthermore, in Table 7, we report additional results by distinguishing two groups of children based on their mother's education – those whose mother attained a high school qualification, first degree, or diploma, and all the other children (mothers with A-level, GCSE/O-level, or below). We do find evidence of heterogeneous production functions in the development outcomes across children of mothers with different educational qualifications. The early language disadvantage is greater for children whose mothers have a higher educational qualification, but is insignificant by age five. In contrast, the negative association between bilingualism and language development is still evident at age five for children whose mothers have an A-level or lower qualifications; however, the effect size is similar in magnitude for children whose mothers have an A-level or lower qualification. Surprisingly, at the age of seven, a greater positive bilingual effect is evident for children of mothers having a low education. This may be driven by selection, i.e., mothers less educated and less proficient in the English language are more likely to speak a foreign language, and they may invest more heavily in home activities to enhance the learning environment of their children (and possibly try to compensate the early bilingual cognitive disadvantage). We cannot disentangle this explanation since our data do not allow us to pin down the language proficiency level.

5.4 Differences in early parental behavior and home environment

Motivated by our findings that bilingual differences in language assessments appear as early as the age of three and that the gap narrows over time, we investigate whether there are corresponding patterns of disparities in maternal behavior and home environment that can explain these early deficits in child development. We hypothesize that potential improvement in children's language development could arise from differences in parent-child interactions, and that a relatively high level of maternal education and time spent with the child has a positive effect on language development. In addition, it is well established that adverse circumstances such as poverty may create a stressful home environment, which contributes to psychological distress in parents and children (Holmes and Kiernan, 2013). This distress may filter into

Table 7 Heterogeneous effects of bilingualism, by mother's level of education

	Age 3	Age 5	Age 7	Age 11	Age 14
<i>Panel (A): Outcome: z-Verbal reasoning skills</i>					
Mother's education – “High”					
Bilingual _{w-1}	-0.430*** (0.152)	-0.205 (0.132)	-0.132 (0.126)	-0.061 (0.147)	-0.270 (0.223)
<i>N</i>	339	310	286	267	241
<i>R</i> ²	0.365	0.318	0.248	0.194	0.248
Mother's education – A-level and below					
Bilingual _{w-1}	-0.328*** (0.067)	-0.200*** (0.071)	0.279*** (0.078)	-0.018 (0.079)	0.019 (0.063)
<i>N</i>	1654	1372	1407	1237	1120
<i>R</i> ²	0.376	0.414	0.217	0.22	0.118
<i>Panel (B): Outcome: z-SDQ score</i>					
Mother's education – “High”					
Bilingual _{w-1}	-0.025 (0.144)	-0.080 (0.116)	0.083 (0.101)	0.100 (0.126)	0.023 (0.188)
<i>N</i>	339	310	286	267	241
<i>R</i> ²	0.199	0.455	0.45	0.459	0.248
Mother's education – A-level and below					
Bilingual _{w-1}	0.128* (0.076)	-0.105 (0.071)	-0.056 (0.063)	-0.135** (0.068)	0.065 (0.075)
<i>N</i>	1654	1372	1407	1237	1120
<i>R</i> ²	0.161	0.343	0.404	0.377	0.276
<i>Panel (C): Outcome: z-Pro-social score</i>					
Mother's education – “High”					
Bilingual _{w-1}	0.404** (0.187)	0.016 (0.146)	-0.022 (0.127)	-0.059 (0.148)	-0.216 (0.191)
<i>N</i>	339	310	286	267	241
<i>R</i> ²	0.121	0.27	0.424	0.382	0.389
Mother's education – A-level and below					
Bilingual _{w-1}	-0.177** (0.078)	-0.006 (0.092)	0.044 (0.074)	0.084 (0.086)	0.110 (0.067)
<i>N</i>	1654	1372	1407	1237	1120
<i>R</i> ²	0.069	0.143	0.227	0.171	0.221

Notes: Standard errors, which are clustered at the family-level, are placed in parentheses. The specifications include the same controls as in Panel (C) in Tables 2–4. Verbal Reasoning skills refer to language development and the SDQ score refers to emotional development.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

parenting practices, which tend to be punitive, inconsistent, and less nurturing. For example, persistent maternal mental health problems have been most strongly associated with child behavioral issues (Noonan et al., 2018). A leading potential explanation in our analysis is that,

in general, immigrant parents foster higher growth rates in learning skills relative to native parents, which could result from different attitudes or parenting behaviors (see Hull and Norris, 2018).

For this part of the analysis, we use dependent variables that indicate the mother's home activities and well-being behaviors: (i) the time she spends with the child; (ii) the frequency of reading to the child; (iii) whether she is in a good relationship with her partner; (iv) whether a doctor has ever diagnosed her with depression or anxiety; (v) her life satisfaction; and (vi) whether she is currently at work.³²

In Table 8, we report our estimates of bilingual maternal behavior toward children from age three to fourteen. We find evidence that mothers who were identified as bilinguals are more likely to spend time with their child; they are more likely to be in a good relationship with their partner and to have higher life satisfaction. They are also less likely to experience depressive symptoms. Overall, the evidence on healthy maternal investments and home environment suggest that children of bilingual families may allow them to “catch up” and develop comparative levels of skills relative to monolingual peers, which is in line with our findings.

5.5 Robustness

The differences between bilingual and monolingual samples, found in Table 1, highlight the challenge which is involved in deriving an empirical method that effectively controls for unobserved child characteristics. Given that we have no experimental design, it is crucial to show that our results are robust to different sample restrictions. Firstly, we expect children whose parents are both foreign to be exposed to foreign languages in a more consistent way than are those with at least one native-born parent. Given this, we restrict the sample to children with two foreign-born parents. For instance, at age three we exclude 607 children on account of having a native-born parent. Table A2 in Appendix shows that results classified on the basis of age group and accounting for child characteristics and past endowments, are robust to this sample restriction. The estimated bilingual effect is very similar to those reported previously in Panel (C) in Tables 2–4, and maintains the result that language disadvantage fades with age. The differences in the SDQ and pro-social outcomes are not significantly different from our main results. The coefficient on past language development is in the range of ≈ 0.44 – 0.19 , indicating that there is a significant loss of language skills over time.

Secondly, we exclude children whose parents arrived in the UK between 1997 and 2002 (this results in 585 children being dropped from the sample). The estimates, which are reported in Table A3 in Appendix, generally show similar results. We also test the robustness of our sample restriction by re-estimating our model on an expanded sample that includes those

³² Specifically, as part of self-completion module, the main respondent (usually biological mother) is asked about the quantum of time she usually spends with the child, and the possible responses are: “plenty of time”, “just enough”, “not quite enough”, and “nowhere near enough”, the higher scores corresponding to a lower quantum of time spent with the child. The variable “frequency reading” is measured on a 5-point scale, with a score of 5 indicating that the mother reads every day and a score of 1 corresponding to “not reading at all”. The variable “good relationship with partner” takes on the value 1 if the mother were to “strongly agree” or “agree” on the question, “Your partner is usually sensitive to and aware of your needs. Do you agree?”, and is asked in waves 1, 2, and 3. In each wave, respondents were asked whether a doctor had ever diagnosed them with depression or anxiety. If they respond positively, the variable “depressive symptoms” is assigned a 1; otherwise the value is 0. The life satisfaction is measured on an 11-point scale, with a higher score corresponding to a better outcome.

Table 8 Bilingual mother's behavior, well-being, and home inputs, by child's age

	(1)	(2)	(3)	(4)	(5)	(6)
	Time spent with child	Frequency of reading	Good relationship with partner	Depressive symptoms	Life	Mother
Age 9 months						
Bilingual mother _{w-1}			0.082** (0.032)	-0.069*** (0.025)	0.231* (0.133)	-0.041 (0.032)
<i>N</i>			3035	3035	3035	3035
Age 3						
Bilingual mother _{w-1}	0.074*** (0.028)	0.174*** (0.067)	0.093** (0.041)	-0.112*** (0.030)	0.329** (0.151)	-0.051 (0.034)
<i>N</i>	1993	1993	1993	1993	1993	1993
Age 5						
Bilingual mother _{w-1}	0.097*** (0.035)	0.039 (0.069)	0.085** (0.039)	-0.053*** (0.019)	0.501*** (0.163)	-0.061* (0.034)
<i>N</i>	1682	1682	1682	1682	1682	1682
Age 7						
Bilingual mother _{w-1}	0.074** (0.032)	0.318*** (0.115)		-0.001 (0.015)	0.265* (0.138)	-0.053 (0.034)
<i>N</i>	1693	1693		1693	1693	1693
Age 11						
Bilingual mother _{w-1}	0.024 (0.044)			0.029 (0.021)	0.267* (0.152)	-0.168*** (0.029)
<i>N</i>	1504			1504	1504	1504
Age 14						
Bilingual mother _{w-1}				-0.021 (0.031)	0.359* (0.186)	-0.191*** (0.031)
<i>N</i>				1361	1361	1361

Notes: Robust standard errors, which are clustered at the family level, are placed in parentheses. All regressions are weighted to account for the non-response rate. Samples have been restricted to children in relation to whom complete information is available on language, emotional, and pro-social behavior. All specifications control for series of mother cohort dummies, ethnicity, marital status, highest grade completed at child's birth, the question of whether mother comes from non-English speaking background, gender of child, number of siblings, poverty indicator, and regional controls.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

who report speaking Welsh ($n = 196$) or Gaelic ($n = 2$) in wave 1. This allows us to consider a notion of bilingualism that is not based solely on the language spoken at home. The estimates, available on request, are qualitatively similar. If anything, the magnitude of the estimates increases slightly in absolute value after the inclusion of the Welsh speakers.

Another potential concern is that any misclassification error induced by the categorical assignment of the language spoken at home may potentially cause a bias in our estimates.³³

³³ See Dustmann and Van Soest (2004) for an analysis of speaking fluency of immigrants using ordered response models with classification errors.

We address this by testing the robustness of our results with an alternative definition of being bilingual. Specifically, we allow the following categories to enter the model: i) mostly English and sometimes other languages spoken at home; ii) only other languages spoken at home, where the omitted category would be placed; and iii) only English language spoken at home. The OLS estimates, available on request, lend overall support to our main findings. Specifically, the levels of language development in children who mostly speak English, but also other languages, at home and those who speak only other languages at home, are well behind the corresponding levels for their “English only” counterparts when they are aged three and five. The negative language impact is greater for children who mostly speak other than English languages at home. By the ages of 11 and 14, the differences in their development are not statistically significant.

We further consider a propensity score matching (PSM) analysis.³⁴ This approach consists of matching “treated” children (i.e., bilingual children) with “untreated” children, based on their observed pretreatment characteristics, and then comparing their language, emotional, and pro-social behaviors. An average treatment effect on the treated (ATT) is obtained by averaging individual-level differences in behavior between the treated and untreated groups. In general, there are several advantages to using PSM methods. First, matching estimators do not impose functional form restrictions, nor do they assume that the treatment effect is homogeneous across the population (Zhao, 2005). Second, with a sufficient vector of observables, matching has been shown to yield estimates that compare favorably with experimental studies (Michalopoulos et al., 2004 and Smith and Todd, 2001). Finally, within the context of our study, a benefit of this approach is that it does not rely on children switching their bilingual status for identification, as in our dynamic analysis. Since it is not necessarily clear a priori which matching algorithm should be implemented, it is a standard practice to present results from multiple techniques (see Anderson, 2013). We consider the multivariate nearest neighbor, kernel matching, and inverse probability weighting (IPW). Technically, the IPW approach, for example, increases the similarity of the distribution of covariates in the treated and control groups via reweighting, which leads to reduced reliance on the functional form of the OLS specification. The multivariate nearest neighbor allows exercise of control over prior achievement.

The PSM estimates, performed by age group, are generally consistent with those reported from OLS models, Panel (B) (without past endowments), with slightly larger effects, on average, being reported from the PSM estimator (see Table A4 in Appendix). In most cases, where the OLS estimates are statistically significant, the corresponding PSM estimates are typically insignificantly different, suggesting that the degree to which PSM tightens the OLS bound is small, and that OLS seems to perform a reasonable job despite our reservations. Regardless of the matching algorithm used, results are consistent with our main findings.

To examine the potential role of unobserved heterogeneity, we also use recently developed tests that explore the stability of the coefficients of interest while increasing the set of control variables. We report estimates of the parameter δ , developed by Oster (2019), that indicates the level of selection on unobserved variables, as a proportion of the level of selection on observed variables, required to drive the treatment effect to zero. The assumptions underlying the calculation of δ can be varied. In our case, we vary the assumed value of $R - max$ and the R^2

³⁴ Propensity score matching may be thought of as assuming the selection problem away because it relies on conditional independence that implies no selection of the unobservable conditional on the observable.

from a hypothetical regression of the outcomes of interest on treatment (bilingual status) and both observed and unobserved controls. If language, emotional, and pro-social outcomes are fully explained by the language spoken at home and full controls are set in the specification, then $R - max = 1$.³⁵

The estimates of δ , which are reported in Table A5 in Appendix, correspond to the bilingual coefficients which are estimated with full controls and past endowments. Results from these tests all exceed 1 in absolute value, aside from the estimates for emotional development at age five.

5.6 Attrition

One potential concern is possible non-random attrition in the MCS, between waves. In Table A6 in Appendix, we compare selective descriptive statistics for individuals who attrite from the sample for any of the three outcomes and those who remain in the sample in wave 6 (age 14). In general, the verbal reasoning scores of children who attrite are slightly lower on average than those of children who do not. Even if attrition is not random, it is important to note that if attrition were attributable to fixed individual characteristics, our use of a first-difference estimator in the GMM implies that our estimates remain unbiased.

We also test whether attrition in any of the three outcomes is random using the approach of Fitzgerald et al. (1998). This approach is based on the assumption that all determinants of attrition can be controlled for (when selection is based on observables). Specifically, we implement a probit model where our dependent variable takes the value of 1 for individuals who drop out of the sample in wave 6 due to non-response in each of the three outcomes, and the value of zero for individuals who remain in the sample. Results reported in Table A7 in Appendix correspond to three separate probit models for the verbal, emotional, and pro-social development on the attrition. The Pseudo *R-squared* results from our attrition probits suggest that baseline variables explain less than 3% of attrition in verbal reasoning, emotional, and pro-social development outcomes between waves 2 and 6.³⁶ Variables that significantly predict attrition in verbal test outcomes include child's past verbal test score, age, and marital status and education of the mother. A Wald test of whether the explanatory variables are jointly equal to zero results in $c^2 = 68.6$, suggesting the joint significance of these variables in predicting attrition. Note that bilingual status and father's/mother's background are insignificant predictors.

We then use the inverse of the fitted probabilities from these probit models to construct weights and adjust our main dynamic estimates, as reported in Table 5. The results, available on request, show that the inverse probability weighted estimates are numerically similar, and qualitatively identical, to the original estimates; therefore, we conclude that the attrition in the main outcomes is not likely to affect our estimates. Since our weights take into account important observable information, including baseline cognitive scores, we conclude that it is unlikely for unobservable factors to drive an attrition process which may substantially change our results.

35 The suggested cut-off to define an "acceptable" level of selection is an estimate of δ (calculated using $(R - max = 1.3 \cdot \bar{R})$) that exceeds 1, with a higher value indicating increased robustness to selection on unobservables (see Oster, 2019 for details).

36 The results are available on request.

Finally, in Appendix, we provide quantitative evidence on the importance of missing data and ensure that this unavailability is, in our case, unlikely to cause bias. We test whether attrition in any of the three outcomes is random using the approach of Fitzgerald et al. (1998) and show that attrition is not likely to affect our results.

6 Conclusion

Relatively few empirical studies evaluate the key issues which surround the development of language in bilingual children over time, despite the availability of burgeoning literature on child cognitive and non-cognitive development, more generally. In this article, we model the language, emotional, and pro-social development and production functions for bilingual children, using cumulative value-added specifications which account for parental investments and children's own ability.

We emphasize several aspects of our findings. First, analysis by child age confirms that very young bilingual children initially have worse language skills than their monolingual peers. The start of schooling appears to attenuate these differences, and at age seven bilingual children show a language advantage. Across the whole age range, we find that, on average, there is no bilingual gap in language development once we control for lagged developmental outcomes and for a rich set of controls in a dynamic panel data framework.

There are several possible channels for this initial result. For example, Feng et al. (2014) show that the bilingual gap in early reading scores is explained by family differences in socio-economic disadvantage as well as in parental, home, and school investment. Both may be particularly important if we consider the improvement in language skills at age seven, and the declining influence of the lag of bilingual status, as being a result of the increasing exposure of bilingual children to English as they age through participation in various social activities, childcare, and starting school. We show that there are corresponding patterns of disparities in parental behavior and home environment that bilingual mothers provide to their children. Specifically, maternal time inputs are greater for children whose mothers are identified as bilingual.

At the same time, our analysis suggests that bilingual boys have fewer behavioral difficulties on average as compared with their monolingual peers. Specifically, all else being equal, speaking a foreign language at home when the boy is seven years old reduces the total difficulty score by 0.19 standard deviations. The outcome persistence parameters are generally low but slightly higher in the emotional development production function; this suggests that the production functions for language and emotional development are different.

We also find a fair amount of heterogeneity along mother's education. For instance, the language disadvantage for children whose mothers are highly educated is found insignificant at the age of five onward; this suggests that parental education does play an important role in children's learning prospects. In comparison, five-year-old bilingual children of mothers with low education continue to fall behind in their expressive verbal skills. Interestingly, at the age of seven, children whose mothers are educated up to the A-level or below mainly drive the bilingual language advantage; it might be the case that these parents start investing more heavily in home activities to enhance the learning environment of their children (possibly trying to compensate the early bilingual cognitive disadvantage).

Ultimately, the most important finding is that, overall, bilingual children are not significantly different when compared with their monolingual peers in language development, taken together with the positive effect of bilingualism on emotional development. With the advent of arguments promoting language acquisition in school from a relatively young age as a means to promote economic competitiveness and growth, the lack of any gap in language, emotional, and pro-social development in bilingual children is vital.

Declarations

Availability of data and material

Data from the MCS is made available by the Centre for Longitudinal Studies (CLS) through the Data Archive. Neither the CLS nor the Data Archive bear any responsibility for the analysis or interpretation of the data reported here.

Ethical standards

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

The research was supported by the Australian Research Council through a Discovery Program Grant (grant no. DP140102491) and the Centre of Excellence for Children and Families over the Life Course (project no. CE140100027).

Author contributions

All authors contributed equally to this manuscript.

Acknowledgments

We would like to thank the following for helpful comments: James Banks, Craig Riddell, Daniel Hamermesh, Ian Walker, Marian Vidal-Fernandez, Melisa Bubonya, and Ivan Stanev, as well as the participants of the annual conference of the Royal Economic Society 2017, the WISERD seminar series at Swansea University, the 6th IZA/SOLE Transatlantic Meeting of Labor Economics 2017, and the RAND 2017 conference.

References

- All-Party Parliamentary Literacy Group Commission, Great Britain; Gavin Barwell; et al. (2012): Boys' Reading Commission: The Report of the All-Party Parliamentary Literacy Group Commission. National Literacy Trust.
- Anderson, D. Mark (2013): The Impact of HIV Education on Behavior Among Youths: A Propensity Score Matching Approach. *Contemporary Economic Policy* 31(3), 503–527.
- Andrabi, Tahir; Jishnu Das; Asim Ijaz Khwaja; Tristan Zajonc (2011): Do Value-Added Estimates Add Value? Accounting for Learning Dynamics. *American Economic Journal: Applied Economics* 29–54.
- Arellano, M.; O. Bover (1995): Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68, 29–51.
- Arellano, Manuel; Stephen Bond (1991): Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies* 58(2), 277–297.
- August, D.; Hakuta, K. (Eds.) (1997): Improving schooling for language-minority students: A research agenda. Washington, DC: National Academy Press.
- Bhatia, Tej K.; William C. Ritchie (2008): *The Handbook of Bilingualism*. John Wiley & Sons (UK).
- Bialystok, Ellen (2006): Second-Language Acquisition and Bilingualism at an Early Age and the Impact on Early Cognitive Development. *Encyclopedia on Early Childhood Development*, 1–4.
- Bialystok, Ellen (2001): *Bilingualism in Development: Language, Literacy, and Cognition*. New York: Cambridge University Press.
- Bialystok, Ellen; Fergus I. M. Craik; Gigi Luk (2012): Bilingualism: Consequences for Mind and Brain. *Trends in Cognitive Sciences* 16(4), 240–250.
- Bleakley, Hoyt; Aimee Chin (2004): Language Skills and Earnings: Evidence from Childhood Immigrants. *Review of Economics and Statistics* 86(2), 481–496.

- Bleakley, Hoyt; Aimee Chin (2008): What Holds Back the Second Generation? The Intergenerational Transmission of Language Human Capital Among Immigrants. *Journal of Human Resources* 43(2), 267–298.
- Blundell, Richard; Stephen Bond (1998): Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics* 87(1), 115–143.
- Bradshaw, Jenny; Rob Ager; Bethan Burge; R Wheeler Pisa (2009): Achievement of 15-Year-Olds in Wales. *Slough: NFER*, 2010.
- Brind, Tom; Caroline Harper; Karen Moore (2008): Education for Migrant, Minority, and Marginalised Children in Europe. CiteSeer.
- Byram, Michael (2004): *Routledge Encyclopedia of Language Teaching and Learning*. Psychology Press, Routledge Taylor & Francis Group, London and New York.
- Caprara, Gian Vittorio; Bernadette Paula Luengo Kanacri; Maria Gerbino; Antonio Zuffiano; Guido Alessandri; Giovanni Vecchio; Eva Caprara; Concetta Pastorelli; Beatrice Bridglall (2014): Positive Effects of Promoting Prosocial Behavior in Early Adolescence: Evidence from a School-Based Intervention. *International Journal of Behavioral Development* 38(4), 386–396.
- Chen, Lijiang; Xia Mao; Yuli Xue; Lee Lung Cheng (2012): Speech Emotion Recognition: Features and Classification Models. *Digital Signal Processing* 22(6), 1154–1160.
- Chin, Aimee; N. Meltem Daysal; Scott A. Imberman (2013): Impact of Bilingual Education Programs on Limited English Proficient Students and Their Peers: Regression Discontinuity Evidence from Texas. *Journal of Public Economics* 107, 63–78.
- Chiswick, Barry R; Yew Liang Lee; Paul W. Miller (2005): Immigrant Earnings: A Longitudinal Analysis. *Review of Income and Wealth* 51(4), 485–503.
- Clifton-Sprigg, Joanna (2016): Best of Both Worlds? Early Cognitive and Non-Cognitive Development of Bilingual Children. Unpublished Working Paper.
- Collier, Virginia P (1989): How Long? A Synthesis of Research on Academic Achievement in A Second Language. *TESOL Quarterly* 23(3), 509–531.
- Collins, Brian A.; Caudio O. Toppelberg; Carola Suárez-Orozco; Erin O'Connor; Alfonso Nieto-Castañón (2011): Cross-Sectional Associations of Spanish and English Competence and Well-Being in Latino Children of Immigrants in Kindergarden, *International Journal of the Sociology of Language* 208, 5–23.
- Costa, Albert; Mireia Hernández; Nuria Sebastián-Gallés (2008): Bilingualism Aids Conflict Resolution: Evidence from the ant Task. *Cognition* 106(1), 59–86.
- Crosnoe, Robert (2007): Early Child Care and the School Readiness of Children from Mexican Immigrant Families. *International Migration Review* 41(1), 152–181.
- Cunha, Flavio; James J. Heckman (2008): Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources* 43(4), 738–782.
- Cunha, Flavio; James J. Heckman; Lance Lochner; Dimitriy V. Masterov (2006): Interpreting the Evidence on Life Cycle Skill Formation. *Handbook of the Economics of Education* 1, 697–812.
- Cunha, Flavio; James J. Heckman; Susanne M. Schennach (2010): Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica* 78(3), 883–931.
- Darcy, Natalie T. (1953): A Review of the Literature on the Effects of Bilingualism Upon the Measurement of Intelligence. *The Pedagogical Seminary and Journal of Genetic Psychology* 82(1), 21–57.
- De Feyter, Jessica Johnson; Adam Winsler (2009): The Early Developmental Competencies and School Readiness of Low-Income, Immigrant Children: Influences of Generation, Race/ethnicity, and National Origins. *Early Childhood Research Quarterly* 24(4), 411–431.
- De Haan, Monique; Erik Plug; José Rosero (2014): Birth Order and Human Capital Development Evidence from Ecuador. *Journal of Human Resources* 49(2), 359–392.
- Del Boca, Daniela; Chiara Monfardini; Cheti Nicoletti (2017): Parental and Child Time Investments and the Cognitive Development of Adolescents. *Journal of Labor Economics* 35(2), 565–608.
- Del Boca, Daniela; Daniela Piazzalunga; Chiara Pronzato (2018): The Role of Grandparenting in Early Childcare and Child Outcomes. *Review of Economics of the Household* 16(2), 477–512.
- Del Bono, Emilia; Marco Francesconi; Yvonne Kelly; Amanda Sacker (2016): Early Maternal Time Investment and Early Child Outcomes. *The Economic Journal* 126(596), F96–F135.
- Department for Education (2013): EAL pupils in schools, The latest statistics about EAL learners in our schools, National Association for Language Development in the Curriculum.
- Department for Work and Pension (2018): Households Below Average Income: 1994/95 to 2016/17.
- Dickerson, Andrew; Gurleen K. Popli (2016): Persistent Poverty and Children's Cognitive Development: Evidence from the UK Millennium Cohort Study. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179(2), 535–558.

- Dustmann, Christian; Najma Rajah; Arthur Van Soest (2003): Class Size, Education, and Wages. *The Economic Journal* 113(485), F99–F120.
- Dustmann; Christian; Arthur Van Soest (2004): An Analysis of Speaking Fluency of Immigrants Using Ordered Response Models with Classification Errors. *Journal of Business & Economic Statistics* 22(3), 312–321.
- Ermisch, John; Chiara Pronzato (2008): Intra-Household Allocation of Resources: Inferences from Non-Resident Fathers' Child Support Payments. *The Economic Journal* 118(527), 347–362.
- Eurodiaconia (2016): The Integration of Children and Families with A Migration Background. An Overview of Projects Among Eurodiaconia Members, European Community Programme for Employment and Social Solidarity.
- Feng, Li; Yunwei Gai; Xiaoning Chen (2014): Family Learning Environment and Early Literacy: A Comparison of Bilingual and Monolingual Children. *Economics of Education Review* 39, 110–130.
- Fiorini, Mario; Michael P. Keane (2014): How the Allocation of Children's Time Affects Cognitive and Noncognitive Development. *Journal of Labor Economics* 32(4), 787–836.
- Fitzgerald, John; Peter Gottschalk; Robert Moffitt (1998): An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics, *Journal of Human Resources* 33(2): 251–299.
- Forrest, Claire L.; Jenny L. Gibson; Sarah L. Halligan; Michelle C. St. Clair (2018): A Longitudinal Analysis of Early Language Difficulty and Peer Problems on Later Emotional Difficulties in Adolescence: Evidence from the Millennium Cohort Study. *Autism & Developmental Language Impairments* 3, 239–270.
- Fromkin, Victoria; Robert Rodman; Nina Hyams (2018): An introduction to language. Cengage Learning.
- Genesee Fred (1989): Early Bilingual Development: One Language or Two? *Journal of Child Language* 16(1), 161–179.
- Genesee, F.; Paradis, J.; Crago, M. (2004): *Dual Language Development and Disorders: A Handbook on Bilingualism and Second Language Learning*. Baltimore, MD: Brookes.
- Goodman Robert (2001): Psychometric Properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry* 40(11), 1337–1345.
- Hakuta, Kenji (1986): Mirror of Language. *The Debate on Bilingualism*. ERIC.
- Han, Wen-Jui (2010): Bilingualism and Socioemotional Well-Being. *Children and Youth Services Review* 32(5), 720–731.
- Han, Wen-Jui; Chien-Chung Huang (2010): The Forgotten Treasure: Bilingualism and Asian Children's Emotional and Behavioral Health. *American Journal of Public Health* 100(5), 831–838.
- Hansen, Kirstine; Elizabeth Jones; Heather Joshi, David Budge (2012): Millennium Cohort Study First, Second, Third and Fourth Surveys: A Guide to the Datasets. London: Centre for Longitudinal Studies, Institute of Education, University of London.
- Hanushek, Eric A.; John F. Kain; Steven G. Rivkin (2009): New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement. *Journal of Labor Economics* 27(3), 349–383.
- Harris, Douglas N. (2010): Education production functions: Concepts. In *International encyclopedia of education*, 3rd ed., edited by Eva Baker, Barry McGaw, and Penelope L. Peterson. Oxford, UK: Elsevier. In press.
- Hastings, Paul D.; Carolyn Zahn-Waxler; JoAnn Robinson; Barbara Usher; Dana Bridges (2000): The Development of Concern for Others in Children with Behavior Problems. *Developmental Psychology* 36(5), 531.
- Heckmann, Friedrich (2008): Education and Migration. Strategies for Integrating Migrant Children in European Schools and Societies. A Synthesis of Research Findings for Policy-Makers; Report Submitted to the European Commission by the Nesses Network of Experts.
- Hoff, Erika (2013): *Language Development*, 5th ed. Belmont, CA: Cengage Learning.
- Holmes, John; Kathleen Kiernan (2013): Persistent Poverty and Children's Development in the Early Years of Childhood. *Policy & Politics* 41(1), 19–42.
- Hull, Marie; Jonathan Norris (2018): The Skill Development of Children of Immigrants. IZA Discussion Paper, IZA DP No. 11724.
- Hundt, Marianne (2006): Bilingual Couples Talk. *The Discursive Construction of Hybridity (Studies in Bilingualism v. 25)*.
- Isphording, Ingo E. (2015): What Drives the Language Proficiency of Immigrants? *IZA World of Labor* 177.
- Isphording, Ingo E.; Sebastian Otten (2014): Linguistic Barriers in the Destination Language Acquisition of Immigrants. *Journal of Economic Behavior & Organization* 105, 30–50.
- Jones, Damon E.; Mark Greenberg; Max Crowley (2015): Early Social-Emotional Functioning and Public Health: The Relationship Between Kindergarten Social Competence and Future Wellness. *American Journal of Public Health* 105(11), 2283–2290.

- Keenan, Kate; Alison E. Hipwell (2005): Preadolescent Clues to Understanding Depression in Girls. *Clinical Child and Family Psychology Review* 8(2), 89–105.
- Kohnert, K. (2010): Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders*, 43(6), 456–473.
- Kovelman, Ioulia; Stephanie A. Baker; Laura-Ann Petitto (2008): Age of First Bilingual Language Exposure as a New Window into Bilingual Reading Development. *Bilingualism: Language and Cognition* 11(2), 203–223.
- Lambert, Wallace E.; John Macnamara (1969): Some Cognitive Consequences of Following a First-Grade Curriculum in a Second Language. *Journal of Educational Psychology* 60(2), 86.
- Lang Kevin; Carlos E. Sepulveda (2007): The Black-White Test Score Differential. Technical Report, Boston University Working Paper.
- Leyendecker, Birgit; Jessica Willard; Alexandru Agache; Julia Jäkel; O. Spiegler; K. Kohl (2014): Learning a Host Country: A Plea to Strengthen Parents' Roles and to Encourage Children's Bilingual Development. *The Challenges of Diaspora Migration: Interdisciplinary Perspectives on Israel and Germany*, 289.
- Locay, Luis; Tracy L. Regan; Arthur M. Diamond Jr. (2013): The Effects of Spanish-Language Background on Completed Schooling and Aptitude Test Scores. *Economic Inquiry* 51(1), 527–562.
- Macnamara, John (1967): Bilingualism and Primary Education: A Study of Irish Experience. *British Journal of Educational Studies* 15(1), 91–92.
- Margetts, Kay (2005): Children's Adjustment to the First Year of Schooling: Indicators of Hyperactivity, Internalising and Externalising Behaviours. *International Journal of Transitions in Childhood* 1, 36–44.
- Michalopoulos, Charles; Howard S. Bloom; Carolyn J. Hill (2004): Can Propensity-Score Methods Match the Findings from A Random Assignment Evaluation of Mandatory Welfare-to-Work Programs? *Review of Economics and Statistics* 86(1), 156–179.
- Millimet, Daniel L.; Ian K. McDonough (2017): Dynamic Panel Data Models with Irregular Spacing: With an Application to Early Childhood Development. *Journal of Applied Econometrics* 32(4), 725–743.
- Mumtaz, Shazia; Glyn Humphreys (2001): The Effects of Bilingualism on Learning to Read English: Evidence from the Contrast Between Urdu-English Bilingual and English Monolingual Children. *Journal of Research in Reading* 24(2), 113–134.
- Najman, Jake M.; Rosemary Aird; William Bor; Michael O'Callaghan; Gail M. Williams; Gregory J. Shuttlewood (2004): The Generational Transmission of Socioeconomic Inequalities in Child Cognitive Development and Emotional Health. *Social Science & Medicine* 58(6), 1147–1158.
- Nantel-Vivier, Amélie; Robert O. Pihl; Sylvana Côté; Richard E. Tremblay (2014): Developmental Association of Prosocial Behaviour with Aggression, Anxiety and Depression from Infancy to Preadolescence. *Journal of Child Psychology and Psychiatry* 55(10), 1135–1144.
- National Center for Education Statistics (2016): Digest of Education Statistics, 2015 (NCES 2016-014).
- Nickell, Stephen (1981): Biases in Dynamic Models with Fixed Effects. *Econometrica: Journal of the Econometric Society* 49(6), 1417–1426.
- Noonan, Katharine; Richéal Burns; Mara Violato (2018): Family Income, Maternal Psychological Distress and Child Socio-Emotional Behaviour: Longitudinal Findings from the UK Millennium Cohort Study. *SSM-Population Health* 4, 280–290.
- Nuffield Foundation (2002): *Languages: the next generation*, The final report and recommendations of the Nuffield Languages Inquiry, London: The Nuffield Foundation.
- Oller, D. Kimbrough; and Rebecca E. Eilers (2002): *Language and Literacy in Bilingual Children*, Vol. 2. *Multilingual Matters*, Sage Publications Ltd (UK).
- Ontario Education Department (2007): *English as a Second Language and English Literacy Development*, The Ontario Curriculum Grades 9 to 12, Ministry of Education.
- Oster, Emily (2019): Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics* 37(2), 187–204.
- Park, K. (2005): *The World Almanac and Book of Facts*. New York, NY: World Almanac Books.
- Plewis, Ian; L. Calderwood; D. Hawkes; G. Hughes; H. Joshi (2007): *Millennium Cohort Study: Technical Report on Sampling*. London: Centre for Longitudinal Study, Institute of Education.
- Portes, Alejandro; Lingxin Hao (2002): The Price of Uniformity: Language, Family and Personality Adjustment in the Immigrant Second Generation. *Ethnic and Racial Studies* 25(6), 889–912.
- Ramrez, N. (2016): Why the Baby Brain can Learn Two Languages at the Same Time. *Recuperado el* 16.
- Raver, C. (2003): Young Children's Emotional Development and School Readiness. *Social Policy Report* 16(3), 3–19.
- Rolfe, Heather; Hilary Metcalf (2009): *Recent Migration into Scotland: The Evidence Base*. Edinburgh: Scottish Government Social Research.

- Rosenthal, Alvin S.; Keith Baker; Alan Ginsburg (1983): The Effect of Language Background on Achievement Level and Learning Among Elementary School Students. *Sociology of Education*.
- Saiz, Albert; Elena Zoido (2005): Listening to What the World Says: Bilingualism and Earnings in the United States. *Review of Economics and Statistics* 87(3), 523–538.
- Sass, Tim R.; Anastasia Semykina; Douglas N. Harris (2014): Value-Added Models and the Measurement of Teacher Productivity. *Economics of Education Review* 38, 9–23.
- Skutnabb-Kangas, Tove (1981): Bilingualism or Not: The Education of Minorities, Vol. 7. *Multilingual Matters*.
- Slavin, Robert E.; Nancy Madden; Margarita Calderón; Anne Chamberlain; Megan Hennessy (2011): Reading and Language Outcomes of A Multiyear Randomized Evaluation of Transitional Bilingual Education. *Educational Evaluation and Policy Analysis* 33(1), 47–58.
- Smith, Jeffrey A.; Petra E. Todd (2001): Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods. *American Economic Review* 91(2), 112–118.
- Spomer, Michelle L.; Emory L. Cowen (2001): A Comparison of the School Mental Health Referral Profiles of Young ESL and English-Speaking Children. *Journal of Community Psychology* 29(1), 69–82.
- Todd, Petra E.; Kenneth I. Wolpin (2003): On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal* 113(485), F3–F33.
- Todd, Petra E.; Kenneth I. Wolpin (2007): The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital* 1(1), 91–136.
- Yang, Sujin; Hwajin Yang; Barbara Lust (2011): Early Childhood Bilingualism Leads to Advances in Executive Attention: Dissociating Culture and Language. *Bilingualism: Language and Cognition* 14(3), 412–422.
- Zhao, Z. (2005): Sensitivity of Propensity Score Methods to the Specifications. Bonn, Germany: Institute for the Study of Labor.

Appendix

Table A1 Children who switched bilingual status over the observed period

	Stayed bilingual		Stayed monolingual		Bilingual – wave 2 and moved to monolingual		Bilingual – waves 2, 3 and moved to monolingual	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Verbal reasoning	6.43	2.39	7.54	2.69	8.16	2.81	6.75	2.78
SDQ index	14.07	3.87	13.10	3.79	12.20	2.66	13.82	4.01
Pro-social index	8.47	1.83	8.42	1.74	8.62	1.66	8.63	1.80
Male	0.41	0.49	0.48	0.50	0.42	0.50	0.49	0.50
White	0.11	0.31	0.82	0.38	0.62	0.49	0.26	0.44
Mother with high degree	0.34	0.48	0.50	0.50	0.69	0.47	0.44	0.50
Father with high degree	0.55	0.50	0.58	0.49	0.76	0.43	0.67	0.48
Below 60% of the median UK income	0.51	0.50	0.19	0.39	0.13	0.34	0.46	0.50
Young migrant	0.49	0.50	0.80	0.40	0.51	0.51	0.53	0.50
Number of siblings	1.89	1.18	1.49	0.98	1.40	1.03	1.89	1.28
Natural mother – in the family	0.97	0.16	0.98	0.13	1.00	0.00	0.98	0.13
Mother non-English	0.91	0.29	0.24	0.42	0.84	0.37	0.84	0.37
Father non-English	0.87	0.34	0.25	0.43	0.80	0.40	0.84	0.37

Table A2 The effects of bilingualism on language, emotional, and pro-social development, excluding native-born parents – by child's age

	Age 3	Age 5	Age 7	Age 11	Age 14
<i>Panel (A): Outcome: z-Verbal reasoning skills</i>					
Bilingual _{w-1}	-0.320*** (0.065)	-0.203*** (0.070)	0.237*** (0.071)	-0.012 (0.074)	-0.087 (0.063)
z-Verbal reasoning skills _{w-1}		0.441*** (0.030)	0.317*** (0.029)	0.290*** (0.037)	0.192*** (0.030)
z-SDQ score _{w-1}		-0.053** (0.025)	-0.143*** (0.028)	0.007 (0.034)	-0.064** (0.031)
<i>N</i>	1386	1148	1191	1058	986
<i>R</i> ²	0.375	0.422	0.225	0.272	0.139
<i>Panel (B): Outcome: z-SDQ score</i>					
Bilingual _{w-1}	0.117 (0.074)	-0.103 (0.069)	-0.052 (0.058)	-0.028 (0.064)	0.008 (0.073)
z-Verbal reasoning skills _{w-1}		-0.049* (0.030)	-0.042 (0.026)	-0.119*** (0.029)	-0.014 (0.038)
z-SDQ score _{w-1}		0.491*** (0.032)	0.506*** (0.027)	0.481*** (0.029)	0.458*** (0.037)
<i>N</i>	1386	1148	1191	1058	986
<i>R</i> ²	0.174	0.349	0.395	0.353	0.250
<i>Panel (C): Outcome: z-Pro-social score</i>					
Bilingual _{w-1}	-0.121 (0.077)	-0.017 (0.089)	0.074 (0.066)	0.043 (0.081)	0.046 (0.064)
z-Pro-social score _{w-1}		0.275*** (0.031)	0.365*** (0.032)	0.270*** (0.041)	0.354*** (0.036)
<i>N</i>	1386	1148	1191	1058	986
<i>R</i> ²	0.059	0.133	0.192	0.171	0.234

Notes: Standard errors, which are clustered at the family level, are placed in parentheses. All regressions are weighted to account for the non-response rate. The results correspond to our full specification, which includes child's age in months, gender, ethnicity and birth weight, mother's age and marital status, mother's and father's highest qualification measured at child's age 9 months, a young migrant indicator, mother and father coming from non-English speaking background, biological mother present in the household, number of siblings, frequency of reading to the child or library visits, poverty indicator, regional controls along with time-varying characteristics measured at period a-1, such as poverty indicator, number of siblings, marital status, presence of biological mother, and lagged input measures for language and emotional development.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

Table A3 The effects of bilingualism on language, emotional, and pro-social development, excluding those who arrived in the UK after 1997 – by child's age

	Age 3	Age 5	Age 7	Age 11	Age 14
<i>Panel (A): Outcome: z-Verbal reasoning skills</i>					
Bilingual _{w-1}	-0.368*** (0.065)	-0.172*** (0.066)	0.215*** (0.073)	-0.040 (0.072)	-0.048 (0.066)
z-Verbal reasoning skills _{w-1}		0.393*** (0.025)	0.298*** (0.027)	0.256*** (0.031)	0.210*** (0.027)
z-SDQ score _{w-1}		-0.035 (0.022)	-0.197*** (0.028)	-0.040 (0.030)	-0.033 (0.029)
<i>N</i>	1812	1525	1491	1329	1194
<i>R</i> ²	0.397	0.402	0.228	0.232	0.141
<i>Panel (B): Outcome: z-SDQ score</i>					
Bilingual _{w-1}	0.109 (0.072)	-0.100 (0.063)	-0.031 (0.056)	-0.117* (0.064)	0.029 (0.075)
z-Verbal reasoning skills _{w-1}		-0.040* (0.024)	-0.067*** (0.023)	-0.100*** (0.025)	-0.022 (0.032)
z-SDQ score _{w-1}		0.502*** (0.026)	0.565*** (0.025)	0.547*** (0.026)	0.520*** (0.033)
<i>N</i>	1812	1525	1491	1329	1194
<i>R</i> ²	0.187	0.365	0.430	0.410	0.309
<i>Panel (C): Outcome: z-Pro-social score</i>					
Bilingual _{w-1}	-0.051 (0.076)	-0.014 (0.079)	-0.004 (0.068)	0.070 (0.077)	0.112 (0.068)
z-Pro-social score _{w-1}		0.311*** (0.027)	0.411*** (0.027)	0.354*** (0.034)	0.399*** (0.034)
<i>N</i>	1812	1525	1491	1329	1194
<i>R</i> ²	0.065	0.153	0.228	0.211	0.266

Notes: Standard errors, which are clustered at the family level, are placed in parentheses. All regressions are weighted to account for the non-response rate. The results correspond to our full specification, which includes child's age in months, gender, ethnicity and birth weight, mother's age and marital status, mother's and father's highest qualification measured at child's age 9 months, a young migrant indicator, mother and father coming from non-English speaking background, biological mother present in the household, number of siblings, frequency of reading to the child or library visits, poverty indicator, regional controls along with time-varying characteristics measured at period $w-1$, such as poverty indicator, number of siblings, marital status, presence of biological mother, and lagged input measures for language and emotional development.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

Table A4 The effects of bilingualism on language, emotional, and pro-social development – PSM analysis

	Age 3	Age 5	Age 7	Age 11	Age 14
<i>Panel (A): Outcome: z-Verbal reasoning skills</i>					
Multivariate nearest neighbor matching					
Bilingual w_{-1}	-0.293*** (0.102)	-0.305*** (0.071)	0.164* (0.092)	-0.089 (0.063)	0.033 (0.071)
Kernel matching					
Bilingual w_{-1}	-0.361** (0.142)	-0.503** (0.215)	0.005 (0.093)	-0.060 (0.125)	0.045 (0.079)
Inverse probability weighting					
Bilingual w_{-1}	-0.374** (0.156)	-0.499*** (0.136)	0.073 (0.129)	-0.024 (0.142)	0.011 (0.066)
<i>N</i>	1993	1682	1693	1504	1361
<i>Panel (B): Outcome: z-SDQ score</i>					
Multivariate nearest neighbor matching					
Bilingual w_{-1}	0.183** (0.093)	-0.012 (0.095)	0.024 (0.082)	0.026 (0.091)	0.018 (0.087)
Kernel matching					
Bilingual w_{-1}	0.106 (0.128)	0.021 (0.152)	0.084 (0.119)	-0.220* (0.125)	0.042 (0.084)
Inverse probability weighting					
Bilingual w_{-1}	0.121 (0.171)	0.078 (0.142)	0.177 (0.198)	-0.208 (0.127)	0.001 (0.083)
<i>N</i>	1993	1682	1693	1504	1361
<i>Panel (C): Outcome: z-Pro-social score</i>					
Multivariate nearest neighbor matching					
Bilingual w_{-1}	-0.228** (0.091)	-0.010 (0.063)	0.065 (0.067)	-0.034 (0.096)	-0.041 (0.070)
Kernel matching					
Bilingual w_{-1}	-0.294** (0.119)	-0.002 (0.214)	0.023 (0.090)	-0.122 (0.142)	-0.073 (0.092)
Inverse probability weighting					
Bilingual w_{-1}	-0.353*** (0.108)	-0.071 (0.159)	-0.051 (0.155)	0.021 (0.264)	-0.062 (0.075)
<i>N</i>	1993	1682	1693	1504	1361

Notes: Bootstrapped standard errors are placed in parentheses.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

Table A5 Delta: size of proportional selection on unobservable to drive “bilingual” to zero

	Age 3	Age 5	Age 7	Age 11	Age 14
<i>Language development</i>					
δ	1.18	1.12	-13.29	1.59	3.31
<i>Emotional development</i>					
δ	1.50	-0.54	1.29	3.25	3.28
<i>Pro-social behavior</i>					
δ	-6.39	-5.79	3.87	6.72	-15.79

Notes: δ is the estimate of delta parameter developed in Oster (2019), which indicates how much selection on unobserved variables would be required to drive the “bilingual” estimate to zero, when measurement is made as a proportion to the selection on observed variables.

Table A6 Baseline summary statistics by attrition by wave 6

Variable	Remained	Attrited	Difference
SDQ	10.17 [5.49]	10.07 [5.56]	-0.10 (0.24)
Verbal reasoning	44.83 [12.52]	43.55 [12.34]	-1.29*** (0.53)
Male	0.47 [0.50]	0.47 [0.50]	0.01 (0.02)
Mother born non-English	0.58 [0.49]	0.54 [0.50]	-0.04* (0.02)
Young migrant	0.6 [0.49]	0.56 [0.50]	-0.04* (0.02)
Mother with higher qualification	0.43 [0.50]	0.37 [0.48]	-0.06*** (0.02)
Father with higher qualification	0.59 [0.49]	0.62 [0.48]	0.03 (0.02)
Below 60% of the median UK income	0.39 [0.49]	0.44 [0.50]	0.05*** (0.02)

Notes: Parentheses contain *t*-statistics. Brackets contain standard deviations.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.

Table A7 Probability models to attrite from the sample on verbal, SDQ and pro-social outcomes

	(1)	(2)	(3)
	Verbal	SDQ	Pro-social
Bilingual	0.024 (0.097)	0.065 (0.095)	0.026 (0.091)
z-Verbal reasoning	-0.078** (0.037)		
z-SDQ		-0.010 (0.029)	
z-Pro-social			0.036 (0.027)
Child's age	0.023** (0.009)	0.023** (0.009)	0.023*** (0.008)
Male	0.062 (0.054)	0.077 (0.054)	0.100** (0.050)
Mother age	-0.011** (0.005)	-0.013** (0.005)	-0.014*** (0.005)
White	0.158* (0.082)	0.170** (0.079)	0.114 (0.075)
Young migrant	0.019 (0.062)	-0.078 (0.061)	-0.033 (0.057)
Mother non-English	-0.040 (0.081)	-0.099 (0.080)	-0.037 (0.078)
Father non-English	-0.079 (0.081)	-0.025 (0.079)	-0.064 (0.078)
Married	-0.147** (0.072)	-0.189*** (0.071)	-0.142** (0.067)
Mother with high degree	-0.187*** (0.067)	-0.213*** (0.066)	-0.207*** (0.063)
Father with high degree	0.008 (0.057)	0.058 (0.057)	0.023 (0.052)
Frequency of reading to the child	0.006 (0.024)	-0.012 (0.025)	0.002 (0.021)
Frequency of the child being taken to the library	-0.004 (0.038)	0.015 (0.037)	-0.007 (0.035)
Below 60% of the median UK income	-0.069 (0.066)	0.020 (0.066)	-0.042 (0.061)
Natural mother presented in Household (HH)	0.292 (0.417)	0.113 (0.476)	0.288 (0.423)
z-Birth weight	0.001 (0.031)	0.033 (0.031)	0.010 (0.028)
Number of siblings	-0.000 (0.026)	0.015 (0.026)	0.010 (0.024)
Wald test	68.6	83.71	77.42
Pseudo R^2	0.0231	0.0278	0.0218

Notes: Parentheses contain t -statistics. Brackets contain standard deviations.

***Significant at the 1% level.

**Significant at the 5% level.

*Significant at the 10% level.