Bilingualism: Language and Cognition

cambridge.org/bil

Greater sensitivity to communication partners' perspectives in children learning a second language at school

Valeria Agostini[1,2], Ian A. Apperly[1,2] and Andrea Krott[1,2]

[1]School of Psychology, University of Birmingham, Birmingham, UK and [2]Centre for Developmental Science, University of Birmingham, UK

Research Article

Cite this article: Agostini, V., Apperly, I.A. and Krott, A. (2025). Greater sensitivity to communication partners' perspectives in children learning a second language at school. *Bilingualism: Language and Cognition*, 1–15 https://doi.org/10.1017/S1366728925000069

Received: 01 April 2024
Revised: 23 November 2024
Accepted: 13 December 2024

Keywords:
second language learning; perspective-taking; communication; Theory of Mind; mentalising; mindreading; bilingual education; foreign language learning

Corresponding author:
Valeria Agostini;
Email: valeria.agostini4@gmail.com

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Abstract

Early learning of a second language at home has been found to be beneficial for children's cognitive development, including their ability to ascribe mental states to others. We investigated whether second language learning in an educational setting can accelerate children's sensitivity to a communication partner's perspective and whether the amount of exposure to second language education makes a difference. We tested three groups of English monolingual four-five year old children with varying language exposure at the beginning of their first year at primary school and 24 weeks later. Children attending bilingual schools and children with weekly second language lessons exhibited similar accelerated development of communicative perspective-taking skills compared to children without second language provision. Such advances were not related to other cognitive advances. Thus, limited foreign language teaching might boost young children's development in communicative perspective-taking skills, providing an enhanced basis for their social competence development.

Research Highlights

- First longitudinal study to suggest that learning a foreign language in the early years of school accelerates sensitivity to communication partner's perspective in monolingual children.
- This research suggests that even minimal exposure to a foreign language in the first year of school can enhance children's social cognition abilities.
- Our findings highlight the potential value of incorporating foreign language teaching into early education curricula to promote the development of social competence.

Speaking more than one language from a very young age has benefits for both individuals and society, including higher academic achievements, greater communicative and intercultural competence, higher salaries, as well as raised employability (see overview in Murphy et al., 2020). Some of these benefits might be driven by advantages in cognitive and social abilities, for instance, the ability to take a communication partner's perspective and to distinguish other perspectives from one's own. This ability is a foundational component of social competence (Devine et al., 2016), that is the skill of building and maintaining relationships with others. Children growing up with an additional language spoken at home (referred to herein as *native* bilingualism) have indeed been found to show a precocious development of perspective-taking skills from infancy into early school years, during the most important years of such development (reviews in Díaz, 2022; Schroeder, 2018). What is not known, though, is whether learning a foreign language as part of a formal education can likewise accelerate communicative perspective-taking development.

1. Introduction

1.1. Precocious bilingual social-communicative development

Living in a bilingual environment and experiencing communicative situations in different languages means that children need to be aware of other people's knowledge and perspectives. This is particularly relevant when communicating with others, suggesting that advantages might be particularly evident in situations of socio-communicative perspective-taking. Indeed, native bilingual children have been shown to be more sensitive than monolingual children to the needs and wishes of communication partners. They adjust their helping behaviour to the needs of communication partners (Gampe et al., 2019), they are better at using referential cues of a speaker during word learning (Yow & Markman, 2011) and are superior at integrating multiple cues in order to understand a speaker's referential intent (Yow & Markman, 2015). Furthermore, they are more sensitive to the knowledge of a communication partner when following the partner's instructions and requests (Fan et al., 2015; Liberman et al., 2017). The latter has been shown with

https://doi.org/10.1017/S1366728925000069 Published online by Cambridge University Press

the means of the director task (Keysar et al., 2000). In this task, a participant and a director are seated at opposite sides of a grid in which various objects are placed. The director asks the participant to move some of the objects. Because some of the grid's compartments are blocked off from the director's view, the participant must take the director's view into account to identify the correct referent objects. Native bilingual children have been found to be more successful in doing so than their monolingual counterparts (Fan et al., 2015). This superior performance appears to develop extremely early in native bilingual children. Evidence for this was presented by Liberman et al. (2017) using a simplified version of the director task with 16-month-old infants. Infants were asked to hand a 'director' one of two identical toys, only one of which was visible to the director. Bilingual infants handed over the mutually visible toy more often than predicted by chance, while monolingual infants performed at chance level. Remarkably, differences are seen after only ~16 months of different sociolinguistic experiences. Given that keeping track of others' perspectives and knowledge is particularly relevant in such socio-communicative situations, the present study focussed on socio-communicative perspective-taking.

1.2. Social-communicative perspective-taking and Theory of Mind

Communicative perspective-taking skills are part of a wider concept, namely the ability to ascribe mental states such as knowledge, beliefs and intentions to other people and to distinguish other people's mental states from one's own, also called Theory of Mind (ToM). Communicative perspective-taking skills can, therefore, be viewed as being a sub-type of ToM. Native bilingual children have been found to outperform their monolingual peers in various types of ToM tasks such as false belief tests, appearance-reality tasks, visual perspective-taking tasks, and socio-communicative perspective-taking tasks (review in Díaz, 2022; Schroeder, 2018). Among these, appearance-reality tasks and false-belief tasks are the most common. For instance, two of the early studies in this field investigated children's understanding of the difference between an object's appearance and its reality, and children's ability to keep these two representations separate (e.g., Bialystok & Senman, 2004; Diaz & Farrar, 2018a, 2018b; Goetz, 2003). Bialystok and Senman (2004) found that bilingual preschool children outperformed monolingual children in appearance-reality tasks, at least once vocabulary knowledge was controlled for. Goetz (2003) found that three- and four-year old Mandarin–English bilinguals performed better on various ToM tasks than Mandarin-speaking and English-speaking monolinguals, including appearance-reality, level 2 perspective-taking and false-belief tasks, even though not all analyses confirmed the general trend. Another common task in this area of research is the classical false-belief task (Diaz & Farrar, 2018a, 2018b; Farhadian et al., 2010; Kovács, 2009; Peristeri et al., 2019), often assessed in a short scenario where one character has superior knowledge over another character, e.g., the change of position of an object or the unexpected content of a container, and the participant needs to keep track of who knows what and how that leads the key character to have a false belief. For instance, Kovács (2009) tested three-year-old bilinguals and monolinguals on a standard false-belief task, a modified false-belief task where one of the characters has a false belief due to their lack of another language, and a control task that tested predictions of events based on physical reasoning. Bilingual children performed better for the two ToM tasks, but not for the control task.

1.3. Impact of foreign language acquisition on perspective-taking skills in educational settings

It is both practically and scientifically important to know whether learning a second language (L2) in a formal educational setting can accelerate the development of communicative perspective-taking skills similarly to growing up bilingually. Practically, any beneficial effects of L2 exposure in an educational setting are potentially deliverable to all, irrespective of family background. Scientifically, variation in educational systems provides unique opportunities to test the effects of different levels of exposure to L2, which is informative for the viability of any practical change in educational practices.

To our knowledge, the impact of learning an L2 in educational settings on the development of communicative perspective-taking has not been studied. It is currently unclear whether full immersion into an L2 is needed to enhance ToM. Children attending a bilingual school have been found not to have increased ToM compared to monolingual children (Buac & Kaushanskaya, 2020). In contrast, children immersed in an L2 preschool have been reported to have enhanced ToM compared to children with five hours of L2 lessons per week (Cheung et al., 2010). One crucial shortcoming of these and other studies on bilingual ToM is that they were cross-sectional and thus did not test participants on ToM or other cognitive abilities before exposure to a second language. The current study therefore tested communicative perspective-taking skills longitudinally, comparing children's skills at the beginning of their L2 learning journey and a few months later.

1.4. Mechanisms of bilingual ToM advantage

It is also important to understand the mechanisms with which L2 learning might accelerate any ToM development. Outside the bilingual literature, individual differences in children's ToM have been related to variations in executive function (EF) skills such as inhibition and cognitive flexibility (see meta-analysis in Devine & Hughes, 2014). It is therefore interesting that native bilinguals have been found to have superior EF skills and evidence suggests that bilingual experiences affect brain networks related to domain-general cognitive control (e.g., DeLuca et al., 2020; DeLuca et al., 2024; Pliatsikas, 2020). Bilingual advantages in EF have been explained with reference to Green's (1998) Inhibitory Control Model of bilingual language processing and Green and Abutalebi's (2013) Adaptive Control Hypothesis. Both models are based on the notion that the two languages of a bilingual speaker are simultaneously activated, meaning that the non-used language, especially a dominant first language, needs to be inhibited. Also, switching between languages is argued to train domain-general cognitive flexibility. Therefore, engaging in a second language is assumed to involve domain-general executive function processes, meaning the bilingual advantage in ToM might be related to an EF advantage (e.g., Bialystok & Senman, 2004; Chan, 2004; Kovács, 2009; Nguyen & Astington, 2014), possibly completely driven by it (Rubio-Fernandez, 2016; Rubio-Fernández, 2017). Similarly, an advantage in communicative perspective-taking due to L2 learning might be caused by a boost in EF skills.

In terms of changes in EF skills due to educational L2 experiences, both full immersion into an L2 and partial immersion into a bilingual school have been found to lead to improved domain-general cognitive flexibility (Carlson & Meltzoff, 2008; Christoffels et al., 2015; Kalia et al., 2019) as well as inhibitory and attentional control (Neveu et al., 2021; Nicolay & Poncelet, 2013; Yang et al.,

https://doi.org/10.1017/S1366728925000069 Published online by Cambridge University Press

2011; but see Lowe et al., 2021). Some studies have found improved EF skills also after short-term L2 learning (e.g., Janus et al., 2016; Rafeekh et al., 2021). Despite this evidence, the overall evidence of a behavioural bilingual EF advantage remains mixed (see recent meta-analyses in Donnelly et al., 2019; Grundy, 2020; Lehtonen et al., 2018; Lowe et al., 2021).

Individual differences in ToM have been related not only to EF but also to variations in language abilities (see meta-analysis in Milligan et al., 2007). Importantly, linguistic abilities have been found to correlate with ToM skills independent of language background, that is for both monolingual and bilingual children (Diaz & Farrar, 2018a; Gordon, 2016). The relevance of language ability for ToM performance is also evident in that bilingual ToM advantages have sometimes been found only after controlling for generally lower language proficiencies (Bialystok & Senman, 2004; Chan, 2004; Diaz & Farrar, 2018a, 2018b; Farhadian et al., 2010; Nguyen & Astington, 2014). Furthermore, general linguistic abilities can be improved by learning a second language. This has been shown in terms of superior metalinguistic awareness of bilingual compared to monolingual children, thus a superior ability to understand language through conscious reflection (see overview in Adesope et al., 2010). Metalinguistic awareness in turn is related to language development more generally. For instance, it predicts vocabulary size in the wider child population (Altman et al., 2018). Furthermore, it has been suggested that exposure to a second language does not only develop second language proficiency but leads to positive transfer to the first language (Lucas et al., 2021) as well as awareness of linguistic features of both the second and first language (Boyd & Ottesjö, 2016; Elvin et al., 2007). Since learning a foreign language, if only for one hour per week and for half a year can boost metalinguistic awareness (Yelland et al., 1993), foreign language learning might lead to a boost in ToM more generally, and communicative perspective-taking skills more specifically, as a consequence of the development of linguistic skills.

In contrast, the level of language skills might not be as relevant to performance on the director task as for other ToM tasks. First, children who were only passively exposed to a second language in Fan et al.'s (2015) study performed very similarly on the director task to children who were active bilinguals. Also, the 16-month-old infants in Liberman et al.'s (2017) study had limited language skills and, given the general tendency of bilingual infants' vocabulary in each language to develop more slowly than monolingual infants' vocabulary (e.g., Bialystok & Senman, 2004; Carlson & Meltzoff, 2008; Chan, 2004; Diaz & Farrar, 2018a, 2018b; Farhadian et al., 2010; Nguyen & Astington, 2014; Nicoladis & Genesee, 1997), it could be argued that the language skills of bilingual infants might have rather been worse than those of their monolingual peers. Nevertheless, bilingual infants performed better in the director task.

To conclude, it is unclear whether the bilingual advantage on the director task might be a consequence of an EF advantage or of superior language skills (see also discussion in Diaz, 2022). Likewise, it needs to be tested whether learning an L2 in an educational setting might accelerate the development of ToM in general, and performance on the director task more specifically, through the development of domain-general cognitive control or language skills.

1.5. The present study

We investigated whether L2 learning in an educational setting accelerates monolingual children's development in communicative perspective-taking skills with the means of the director task. Furthermore, it asked whether the amount of exposure is important by testing whether the effect of minimal exposure through a weekly lesson would be comparable to high exposure through immersion into a bilingual school. In contrast to previous bilingual ToM studies, we employed a longitudinal design that assessed children at the beginning of their L2 exposure (beginning of their first year of primary school) and 24 weeks later. All children were English-speaking monolinguals when entering school. We compared the performance of children who attended bilingual schools that taught about ~50% in an L2, those who attended schools with a weekly L2 lesson (L2 learners) and those who attended schools without any L2 provision (control group). If L2 learning in an educational setting accelerates the development of communicative perspective-taking skills, then both bilingual school children and L2 learners should improve their performance on the director task more strongly than those without any L2 provision. A comparison between the two L2 learning groups shows whether the amount of L2 educational exposure affects the development of communicative perspective-taking skills. Furthermore, we investigated whether any improvements in such skills were an indirect consequence of improvements in EF or language, measuring children's EF skills (conflict processing and cognitive flexibility) and language skills (English vocabulary) at both testing points.

2. Methods

2.1. Participants

We estimated the minimum number of participants on the basis of two main power analyses with follow-up analyses, using the Shiny app by Lakens and Caldwell (2021; see also https://shiny.ieis.tue.nl/anova_power/). In the first analysis, we estimated the sample size for a Group × Time interaction for performance on the director task. The second analysis estimated the sample size for finding a significant Group effect as well as pairwise group differences at the second testing point (T2). We based both analyses on the results of Fan et al. (2015), who compared performance on the director task of three similar groups of children as in our study, namely monolingual children, native bilingual children, and children passively exposed to a second language. For the within-between 3 (Group) × 2 (Time) ANOVA, we assumed that all three participant groups would perform like the monolingual participants in Fan et al.'s (2015) study at the beginning of our study (i.e., average 50% correct responses). We also assumed that children attending bilingual schools and weekly language learners, but not controls would increase the level of the performance of children passively exposed to another language in Fan et al.'s (2015) study (i.e., an average of 77% correct responses). Furthermore, we assumed the same standard deviation for responses as in Fan et al.'s study (i.e., 29.4%). Thirty participants in each group, 2000 simulations, a correlation of $r = 0.5$ for scores between the two testing points, and $\alpha = .05$ showed a power of 95.85 and an effect size of Cohen's $d = 0.47$ for a significant Group × Time interaction. Since exposure to a second language in our study was relatively short (six months), the effect might be slightly smaller than that in Fan et al.'s (2015) study. The power dropped to 76.5 and Cohen's $d = 0.35$, when the performance of the two L2 learner groups was set to 70% instead of 77%.

Second, a simulation analysis with 2000 simulations, $\alpha = .05$, and 24 participants in each of the three groups, who performed like the groups in Fan et al. (2015), suggested a power of 90.0 and Cohen's $d = 0.47$ to detect the main effect of the group. A follow-up power analysis comparing a pair of participant groups showed that the

https://doi.org/10.1017/S1366728925000069 Published online by Cambridge University Press

power to reach the same performance difference as observed between monolingual children and native bilingual children in Fan et al.'s (2015) study with 24 participants in each group was 85.8 with Cohen's $d = 0.91$. The power to reach the same performance difference as between the L2 exposure children and monolingual children in Fan et al.'s (2015) study was 88.5 with Cohen's $d = 0.94$. Based on these results, we concluded that 30 participants should be sufficient to detect a significant Group × Time interaction, as well as a significant difference between at least two of the groups. Taking a conservative approach and because we expected attrition of participants, we aimed for 33 participants per group.

We recruited 109 4- to 5-year-old children in their first year at an English primary school who were all reported to be monolingual by their parents (see below for more details of their language exposure background). Of those, children who were exposed to an L2 at home ($n = 6$), were reluctant to complete more than one task ($n = 2$), or had moved school after the first testing point ($n = 2$) were excluded from the analysis. The final sample therefore included data from ninety-nine 4 to 5-year-olds, all monolingual before entering school. Thirty-two children attended bilingual schools (BilS), 29 received weekly L2 classes in mainstream schools (L2 learners) and 38 did not have any L2 instruction (NoL2).

Children in bilingual education were recruited from two schools, one located in Oxfordshire and the other in South East London in the United Kingdom, that taught in an L2 for approximately half of the school time, but with different structures of L2 immersion. In one school, English was used as a medium of instruction for the core subjects (maths, English and science) in the morning, while L2 (French) was used for the remaining subjects (e.g., sports and arts) in the afternoon. In the other school, all subjects were taught in L2 (either French, German or Spanish) for half of the week and in English for the other half. L2 learners were recruited from 4 mainstream schools and received 30- to 60-minute L2 classes once a week (either French, German or Spanish), led by an L2 specialist teacher. Both bilingual school children and L2 learners were taught in person in their classrooms. Since the children were four years old and attended the first year of primary school, they were not yet able to read or write. The teaching mode was therefore oral. In the early years of school in the United Kingdom, foreign language learning is fostered through natural, immersive activities like songs, games, storytelling and role-play, allowing children to acquire language in context (Lightbown & Spada, 2013; Pinter, 2017). The NoL2 control group was recruited from two mainstream schools with no L2 provision. All schools were from neighbourhoods whose ethnicity was predominantly White, followed by Asian and Black (Office for National Statistics, 2021). The majority of the schools were state-funded, apart from two schools with L2 provision, which required parents to pay a fee. Furthermore, all schools followed the national curriculum. Since children attended the first year of primary school, they were taught primarily through games and play. Areas of learning in the first year are communication and language, personal, social and emotional development, physical development literacy, mathematics, understanding the world, and expressive arts and design. Given the young age of the children, a large part of the day is spent in free play, with periods of focused work in small groups where children start to learn, for instance, to read, spell and simple maths. Most relevant to the present study, are activities around language and communication. There is a focus on vocabulary development. This is achieved through reading to the children and engaging them in the stories. Teacher-led one-to-one conversations, with sensitive questioning, invite children to elaborate and use new words in a range of contexts. In addition, children are encouraged to act in role-play. We are not aware of any markable differences in these activities across the participating schools.

Table 1. Participant characteristics by group

	BilS ($n = 32$)	L2 learners ($n = 29$)	NoL2 ($n = 38$)
M age in months (range)	56.0 (50–62)	57.2 (51–64)	59.5 (53–65)
Gender (f/m)	11/21	15/14	21/17
Maternal background UK[1]	96.9%	96.5%	100%
Paternal background UK[2]	93.8%	100%	100%
M SES (0–1) (SD)	.78 (.19)	.71 (.21)	.76 (.16)
M hours of extra-curricular activities/week (SD)	1.2 (.3)	1.6 (.2)	1.4 (.2)
M hours of computer usage/week (SD)	2.2 (.5)	2.4 (.4)	2.6 (.6)
Mdn number of siblings (IQR)	1.0 (1–2)	1.0 (1–2)	1.0 (1–1.25)
Mdn number of older siblings (IQR)	.5 (0–.5)	1.0 (1–1.5)	1.0 (0–1)
M BPVS score (SD)	74.1 (12.6)	70.9 (13.7)	72.6 (14.0)
M IQ (Raven's)	15.4 (3.3)	14.8 (3.7)	15.1 (3.6)

Note.[1]Non-UK places of birth for mothers were USA (n = 1) in the BilS group and Ireland (n = 1) in the WL2 group.
[2]Non-UK places of birth for fathers were New Zealand ($n = 1$) and South Africa ($n = 1$) in the BilS group.

Information on family language background, parents' education and employment, annual family income, children's age and extra-curricular activities was collected through a parental questionnaire (for details see Table 1 and description below). All parents reported that both they and their children were born and brought up in the United Kingdom or in an English-speaking country, that English was the only language spoken at home and that the children had no exposure to any other language than English at home. All children had resided in the United Kingdom since birth except for twins in the BilS group who had lived in Singapore between the ages of 5 weeks and 15 months. But they were, like the other participants, monolingual English speakers. At the beginning of the study, there were no gender differences among the three participant groups, $\chi2(2) = 3.34$, $p = .189$, SES, $F(2,95) = 1.0$, $p = 0.389$, number of hours of extra-curricular activities per week, Kruskal–Wallis[1] $\chi2(2) = 4.0$, $p = .135$, number of hours of computer usage per week, Kruskal–Wallis $\chi2(2) = 1.3$, $p = .511$, the total number of siblings, Kruskal–Wallis $\chi2(2) = 1.0$, $p = .600$, number of older siblings, Kruskal–Wallis $\chi2(2) = 2.1$, $p = .354$, and non-verbal general IQ (Raven's Coloured Progressive Matrices; Raven et al., 1998), $F(2,96) = .2$, $p = .790$. However, the groups differed in age, $F(2,96) = 7.2$, $p = .001$, $\eta_p^2 = .131$ (see participant characteristics in Table 1). NoL2 children were significantly older than BilS children ($p = .001$) and tended to be older than L2 learners ($p = .063$). Given the group differences in age, we checked whether any of our measures of interest correlated with age. This was only the case for the performance improvement of cognitive flexibility, measured

[1]Note that we conducted Kruskal–Wallis tests whenever the assumption of normal distributions was violated.

https://doi.org/10.1017/S1366728925000069 Published online by Cambridge University Press

by the Dimensional Change Card Sort test (DCCS; Zelazo, 2006; Pearson $r(99) = -.252$, $p = .012$; all other $ps > 0.05$). Thus, while no group differences could be explained by age differences, we took age into account when investigating improvements on the DCCS.

2.2 Procedure

Parents provided consent for the participation of their children in the study. They also filled in a background information questionnaire before the start of the study. Children took part in a battery of tests: a child-adapted version of the director Task (Keysar et al., 2000) as a measure of socio-communicative perspective-taking, the Coloured Progressive Matrices (Raven et al., 1998) as a measure of general IQ, the British Pictures Vocabulary Scale III (BPVS; Dunn et al., 2009) as a measure of schools' teaching effectiveness, and the Attention Network Test for children (ANT; Rueda et al., 2004) and the Dimensional Change Card Sort test (DCCS; Zelazo, 2006) as measures of EF. We chose the ANT and the DCCS as two very common tasks measuring inhibition/attention and cognitive flexibility, respectively. Both have been found to be related to ToM in general and in bilingual children in particular (Buac & Kaushanskaya, 2020; Devine & Hughes, 2014). Tests were administered in the first months of the school year when all children had minimal or no exposure to an L2. The tests were repeated after 24 weeks. Both times, children were tested over three testing sessions, taking place on different days over the course of two weeks. In the first session (30 minutes) participants completed Raven's Matrices, followed by a divergent thinking task (results for this task are unrelated to the present research question and are not reported) and the DCCS. In the second testing session (30 minutes), they completed the ANT and the BPVS. In the final session, they took part in the director Task (15 minutes).

The study followed the ethical guidelines of the British Psychological Society and was approved by the Science, Technology, Engineering and Mathematics (STEM) Ethics Committee of the University of Birmingham.

2.3. Materials

2.3.1. Background information questionnaire

A modified version of Luk and Bialystok's (2013) Language and Social Background Questionnaire provided information on participants' characteristics such as gender, date of birth, extra-school activities (type and amount of hours per week), computer and video game usage (amount of hours per week), number of older and younger siblings, maternal and paternal country of birth and upbringing, maternal and paternal level of education, maternal and paternal employment status and occupation, annual family income, participant's place of birth and previous residence in non-English speaking countries, languages spoken by participant and daily exposure to other languages besides English at home. Parents were also asked to describe their child's language status by choosing from four options: monolingual, exposed to different languages, bilingual or multilingual. Only children who were described as monolingual and had no exposure to other languages besides English at home were accepted in the study.

We obtained a socio-economic status (SES) index for each child by averaging indices of parental education, occupational status and annual income. The latter were calculated as follows. The level of education for each parent was measured on a six-point scale, with one being "no formal educational qualification" and six "masters/ Doctoral degree, National Vocational Qualification level 5, or equivalent". The occupational status of each parent was classified according to the Standard Occupational Classification Hierarchy redacted by the Office for National Statistics. Indices for education and occupational status were obtained by averaging the scores of both parents (if applicable). Finally, annual family income was measured on a seven-point scale, with one being 'less than £15,000' and seven '£65,000 or more'. All three indices were converted onto a 0–1 scale before being averaged for the SES index.

2.3.2. Director task

The ability to take the interlocutor's perspective during a conversation was assessed through the director task (Keysar et al., 2000), a widely used test of communicative perspective-taking. Our procedure was adopted from Fan et al. (2015), who found superior performance by native bilingual children compared to monolingual children. Children followed the instructions given by a director and moved objects around a 4 × 4 array of boxes (see example configuration in Figure 1). An adult assistant played the part of the director, while the experimenter took note of the child's behaviour. The assistant and child were seated on opposite sides of the array, facing each other, and the experimenter sat beside the child. On the director's side, 4 boxes were blocked by slats, thus creating a limited view of the objects on the grid. Eight objects were placed on the grid, three of which were in the slatted boxes. To familiarise the children with the task and the differences in visibility of the objects depending on the view, children first played the part of the director, while the assistant responded to their instructions. The experimenter said: "We are playing a game with this grid. I put some objects on the grid, but from this side, we cannot see all the objects that are on the grid. She/He [the assistant] can see everything that is on the grid, but we cannot see everything because some objects are hidden behind these slats [pointing to the blocked boxes]". "First, you tell her/him [the assistant] to move something on the grid, then, if she/he moves the right object, we will swap places, and she/he will tell you what to move". Then, the experimenter suggested the first instruction to the child: "Can you see two cups? One big and one small? Tell her/him to put the big cup next to the green frog". After the child had repeated the instruction, the assistant moved a bigger cup, which was in a blocked box, deliberately making a mistake. At this point, the experimenter, with a tone of confused surprise, would ask the child: "Oh! Did she move the right cup?" If the child did not acknowledge the mistake, then the experimenter said "Yes, that's a big cup but we could not see that big cup from our side!" and explained the error. Afterwards the experimenter said to the assistant: "Let us try again! Remember that from this side we cannot see everything that you can see". A second critical instruction was presented, and this time the assistant selected the target object. The child was then shown the grid from the assistant's privileged point of view and made aware that the correct object had been selected despite the presence of a referential alternative (the same object in a different size) in one of the blocked boxes.

The two familiarisation trials were followed by 12 experimental trials, with the adult assistant taking over the role of the director and the child responding to the instructions. Thus, for the experimental trials, the child participant and the adult assistant swapped places so that the child had the full view of all objects, while the assistant had a limited view (see Figure 1). The trials were presented in four different displays. For each display, three instructions were given, one of which was critical due to the presence of a distractor object in the privileged ground. To reduce learning effects, the grid was rotated 90° anticlockwise after each set of instructions so that the covered slots changed their position in the grid. During the instructions, the director kept his/her gaze towards the center of the grid, in order to not give away

https://doi.org/10.1017/S1366728925000069 Published online by Cambridge University Press

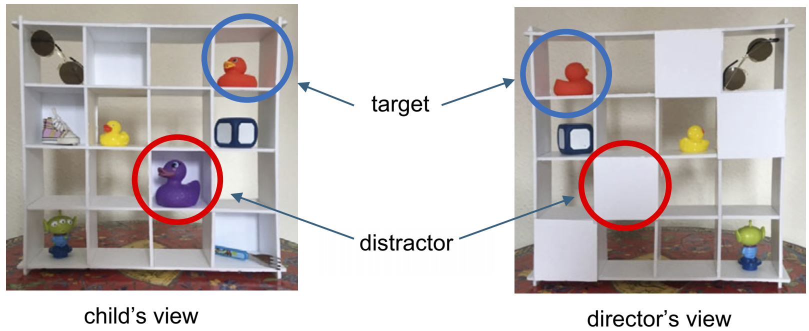

Figure 1. Example experimental setup for the director task and the request to 'move the big duck next to the alien'.

which object needed to be moved. One point was awarded for each correct selection in critical trials, with a maximum score of four points. Scores were calculated on children's first responses, ignoring any repairs. At the second testing point of the study, all objects were replaced with new ones, but the presentation order of critical instructions was kept the same as at the first testing point.

We are not aware of any reliability tests for the director task. We therefore report reliability measures of our study as part of the results section.

2.3.3. Coloured progressive matrices

The coloured progressive matrices (Raven et al., 1998) are a commonly used measure of non-verbal reasoning, also utilised to compare bilingual with monolingual participants on general IQ (Cheung et al., 2010; Janus et al., 2016; Nicolay & Poncelet, 2013; Rafeekh et al., 2021). Children were presented with 36 incomplete figural patterns (sets A, B and AB) and were asked to select the missing part of each pattern from a choice of six options. For each correct answer, one point was awarded, with a maximum score of 36. Stimuli were presented on a laptop via E-prime (E-Studio 2.0). Children pointed to the chosen option. The experimenter entered the choice by pressing the corresponding buttons on the keyboard.

Previous studies have reported excellent test–retest reliability (around .90) and good split-half reliability values ranging from .65 to .94, with most studies reporting values above .80 for children in the early school years (Raven et al., 1998).

2.3.4. The British Pictures Vocabulary Scale III (BPVS)

The BPVS (Dunn et al., 2009) is a widely used standardised measure of receptive vocabulary. Children are shown tables with four pictures and are asked to point to the one that matches a word produced by the experimenter. Tables are grouped into 14 sets of 12 items. Following the guidelines of the test, children worked through the tables until they made eight or more incorrect selections within a set. A score was calculated by subtracting the total number of errors from the number of tables presented. The BPVS III does not report reliability, but its predecessor, BPVS II, has excellent reliability (e.g., for 3–5 year-olds: split-half reliability 0.89, Cronbach's alpha 0.96; Dunn et al., 1997).

2.3.5. Attention Network Test for Children (ANT)

We measured inhibition ability with a child-friendly flanker task, the Attention Network Test for Children (ANT), implementing the version by Rueda et al. (2004). This measure has been widely used to test inhibition ability, including in research on EF in bilingual children (e.g., Buac & Kaushanskaya, 2020; Carlson & Meltzoff, 2008; Neveu et al., 2021; Yang et al., 2011). Participants 'fed' a hungry fish by pressing a left or right button according to the orientation of the fish. In neutral trials, a single yellow fish faced left or right, while in congruent and incongruent trials the target fish appeared in the middle of a row of five identical fish. The four flanking fish either faced the same direction as the target fish (congruent trials) or the opposite direction (incongruent trials). All trials started with a fixation cross for 400 ms in the middle of a laptop screen, followed by a warning cue (an asterisk) for 150 ms and a stimulus for 450 ms. Attentional cues appeared in the center, above or below the position of the fixation cross. Stimuli appeared above or below the position of the fixation cross. Stimuli were displayed until the child responded, with a maximum duration period of 1700 ms. Correct responses received both visual and auditory feedback, as the target fish moved its mouth, making bubbles and a 'woo-hoo' sound was played. Incorrect or late responses were followed by auditory feedback only, namely a 'wrong buzzer' sound. Stimuli were presented on a laptop via E-prime (E-Studio 2.0). Responses were recorded using a Cedrus RB-844 response pad with a right button showing a picture of the target fish oriented to the right and a left button showing the target fish oriented to the left.

The experimenter introduced the task to the children following the procedure described in Rueda et al. (2004). They placed a laminated card with a right-facing target fish above the buttons. Children were told that the fish was very hungry, and their task was to feed him by pressing one of the two buttons. Pointing first to the card and then to the right button, the experimenter said: "Sometimes the fish is facing this way. In this case, you need to press this button". Then, presenting a card with a target fish facing left and, pointing to the left button, the experimenter continued "but sometimes the fish is facing the other way. So, which button do you need to press if he is facing the other way?". The correct answer was suggested if the child showed hesitation. Afterwards, the experimenter presented a card with a row of five fish, all facing left and, pointing to the central one, said: "Sometimes the fish is not alone on the screen, he is with other fish, but you need to look at him in the middle. Which button do you press to feed him?" and, showing the right-oriented row of fish, "Which button do you press if he is facing the other way?". Finally, the experimenter presented a card showing one of the two incongruent flankers and, pointing to the target fish, said: "I told you that you always have to look at the fish in the middle and not at the others because sometimes the other fish can be tricky. They want to trick you and they go in the opposite direction. But you need to feed the fish in the middle and not the others, so which button do you press for the fish in the middle?" and, showing the other flanker stimuli card, continued "And which button do you press for this other one?". Correct responses were demonstrated, and further explanations were given in case of

https://doi.org/10.1017/S1366728925000069 Published online by Cambridge University Press

incorrect answers. Children were told that there would be a small cross in the middle of the screen and that they had to keep their eyes on the cross as the target fish would appear above or underneath it. The presence of different cue types trials was not mentioned. Children were encouraged to keep their right pointer finger on the right button and their left pointer finger on the left button before starting the task and told to press the relevant button as quickly as possible when the fish appeared on the screen.

There were 24 practice trials followed by two blocks of 48 trials, consisting of an equal number of trials in each condition and an equal number of trials with the target fish facing left or right, presented in random order. Accuracy and reaction times (RTs) were recorded. Omissions of response were recorded as errors. We calculated three indices: conflict index, alerting index, and orienting index. For the conflict index, we subtracted children's performance in congruent trials (where no conflicting stimuli had to be resolved) from their performance in incongruent trials (where target fish and flankers were in competition). These two subindices were standardised and summed, after reversing RTs so that higher values of both RT and accuracy represented better performance (see also Liesefeld et al., 2015). For the alertness index, we subtracted the mean of the Double Cue condition of the ANT from the mean of the No Cue condition. We calculated an alerting index by adding reversed standardised RTs and accuracy scores, so that higher values corresponded to higher performance. For the orienting index, we subtracted the mean of the Spatial Cue condition from the mean of the Central Cue condition. We calculated the orienting index by adding reversed standardised RTs and standardised accuracy scores so that higher values corresponded to higher performance.

Child-friendly versions of the ANT have shown excellent reliability in terms of overall response times and errors, good reliability in terms of test–retest reliability for the inhibition effect, but problematic reliability for both altering and orienting effects (e.g., Casagrande et al., 2022; Luna et al., 2021). We therefore need to treat results for the latter two indices with caution.

2.3.6. Dimensional Change Card Sort (DCCS)

As a second EF measure, we administered both the standard and the advanced versions of a measure of children's cognitive flexibility, the DCCS (Zelazo, 2006), which has been widely used to assess EF in bilingual children (e.g., Buac & Kaushanskaya, 2020; Carlson & Meltzoff, 2008; Devine & Hughes, 2014; Fan et al., 2015; Neveu et al., 2021). We created a composite score by summing the scores of the standard and the advanced version of the test. In both versions, children were asked to sort cards according to colour or shape of the depicted objects. At both testing points, the same procedure was followed, but the objects and colours differed (first testing point: red rabbit and blue boat; second testing point: yellow flower and blue car).

In the standard version, children were first asked to sort cards by one dimension (colour) and halfway through by another dimension (shape). Two practice and 12 testing cards (7 × 11 cm) were used. At the first testing point, half of the cards showed a red rabbit while the other half showed a blue boat. The cards needed to be sorted into two open transparent containers (16.8 × 11.6 × 4.5 cm). A target card was attached at the back of each container, clearly visible to the child. The target cards showed a blue rabbit and a red boat so that the cards to be sorted were never identical to the target cards. The experimenter explained to the children that they were going to play some games with the cards. The experimenter said: "The first game is called the *colour game.* In the colour game, you need to sort the cards by colour, so all the red cards go here [pointing to the container featuring the red target card] and all the blue cards go here [pointing to the container featuring the blue target card]". Then, the experimenter showed the red rabbit practice card and said: "Here's a red card. Where does it go?" Children were asked first to point to the correct container and then to place the card face down into it. The same was repeated with the blue boat practice card. Correct responses were demonstrated if the child made mistakes. After the two practice trials, six pre-switch trials were presented, consisting of three red rabbit cards and three blue boat cards in pseudorandom order. No feedback was given. After the child completed the six pre-switch trials, the experimenter said: "Now we are changing the game, we are playing another game called *picture game.* In the picture game, you need to sort cards by picture, so all the rabbits go here [pointing to the container with the target rabbit card] and all the boats go here [pointing to the container with the boat target card]. Here's a rabbit, where does it go?". The card was handed to the child who put it face down in one of the containers, no feedback was given. There were six post-switch trials with three red rabbit cards and three blue boat cards in pseudorandom order. Both in pre- and post-switch trials the card was presented mentioning the relevant feature ("Here's a red/blue card" for pre-switch, "Here's a rabbit/boat card" for post-switch) and handed to the child who put the card face down into one of the containers. One point was given for each card sorted correctly, with a maximum score of 12 points.

Immediately after the child completed the standard version, the advanced version was administered. In this version, the sorting rule was indicated on the cards, with a rainbow for sorting by colour and black outlines of a rabbit and a boat for sorting by shape. A new set of two practice cards and 12 testing cards was used: eight red rabbit cards and six blue boat cards. Half of the rabbit cards and half of the boat cards were colour-sorting cards, the other half were shape-sorting cards. Cards were presented in a pseudorandomised order so that in 50% the rule needed to be switched, while in the other 50%, the rule stayed the same. The first two cards were used for practice.

The experimenter showed the new set of cards to the child and said: "Now we are going to play another game with this special set of cards. In this set, some cards have a rainbow and some cards have two small pictures. If the card has got a rainbow [showing a rainbow-labelled rabbit card], it's a colour-game card and you have to play the colour game, but if the card has got two small pictures [showing an outlines-labelled rabbit card], it's a picture-game card and you need to play the picture game". Then, the experimenter handed the rainbow card to the child and said: "This has got a rainbow, so it's a colour game card. Can you remember the rule for the colour game?" If the child failed to remember the rule, the experimenter said: "In the colour game, red cards go here and blue cards go here [pointing to the corresponding target cards]. This is a colour game card, where does it go?" If the child placed the card in the wrong container, the mistake was corrected, and the rule reminded again. The same procedure was repeated with a picture-game rabbit card. The two practice trials were followed by 12 trials presenting three picture-game rabbit cards, three colour-game rabbit cards, three picture-game boat cards and three colour-game boat cards in a pseudo-random order. In all trials, the card was presented hinting at the relevant rule ("Here's a colour-game/picture-game card"). For each card sorted correctly, one point was rewarded, with a maximum score of 12 points. The maximum total score for both the standard and advanced version was 24 points. Scores were calculated on children's first sorting responses, ignoring any corrections.

Both the standard version and advanced versions of the DCCS have shown excellent test–retest reliability in children (ICCs = .90–.94; Beck et al., 2011).

3. Results

We first investigated differences in the groups' development across T1 and T2 in EF, language, and development of perspective-taking skills by conducting ANOVAs with the factors Time (T1 vs T2) and Group (BilS, L2, NoL2). Significant Time × Group interactions were followed up with paired t-tests to explore improvements of scores across time points for each group as well as one-way ANOVAs at T1 and T2. In case of no significant interactions, but significant main effects of the Group, we conducted one-way ANOVAs to test differences between the groups on combined scores for T1 and T2. We conducted Welch's tests and, if significant, Games-Howell post-hoc tests whenever Levene's test of homogeneity of variance was violated. Second, we then conducted mediation analyses to test whether the development of perspective-taking skills (director task) might be caused by development in EF (DCCS and ANT) or language (BPVS). Descriptive statistics of all measures can be found in Tables S1 and S2 in the Supplementary Material. All data, analyses and materials for the study are available at the Open Science Framework at https://osf.io/s9ghz/?view_only=d3d0bdf8a5a443a581d78b4ff51e746b.

3.1. Dimensional Change Card Sort task (DCCS)

The left panel of Figure 2 shows the results of the DCCS ($n = 99$). The ANOVA showed a significant main effect of Time, $F(1, 96) = 124.7$, $p < .001$, $\eta_p^2 = .565$, 95% CI [.42, .68], and Group, $F(2, 96) = 3.6$, $p = .033$, $\eta_p^2 = .069$, 95% CI [.00, .18], and a Time × Group interaction, $F(2, 96) = 11.1$, $p < .001$, $\eta_p^2 = .188$, 95% CI [.06, .33]. All groups showed an improvement between T1 and T2, BilS: $t(31) = 12.0$, $p < .001$, Cohen's $d = 2.12$, 95% CI [1.49, 2.75], L2: $t(28) = 5.39$, $p < .001$, Cohen's $d = 1.00$, 95% CI [0.55, 1.44], NoL2: $t(37) = 3.4$, $p = .002$, Cohen's $d = 0.55$, 95% CI [0.21, 0.89]. However, while there was no difference between groups at T1, $F(2, 96) = .2$, $p = .807$, $\eta_p^2 < .001$, 95% CI [.00, .04], there was at T2, Welch $F(2, 57.7) = 17.5$, $p < .001$, $\eta_p^2 = .27$, 95% CI [.12, .41]. Games-Howell post-hoc tests for T2 showed higher DCCS scores for BilS children than L2 learners, $t(42.3) = 3.01$, $p = .012$, Cohen's $d = .78$, 95% CI [0.25, 1.29], and NoL2 children, $t(56.5) = 5.49$, $p < .001$, Cohen's $d = 1.38$, 95% CI [0.84, 1.88], while scores for L2 learners and NoL2 children did not differ, $t(60.9) = 1.74$, $p = .161$, Cohen's $d = .46$, 95% CI [−.03, .95]. Thus, BilS children improved more strongly than L2 learners and NoL2 children on the DCCS.

3.2. Attentional Network Test (ANT)

We noticed that some children had very high error rates in the ANT task. We therefore calculated how many correct trials were needed for a participant to score statistically above chance (i.e., 50% correct responses). Making 57 correct responses out of 96 total responses has a one-tailed probability of 0.041 by binomial test. Thus, we included in the analysis of ANT indices only children who made at least 57 correct responses (60% accuracy) at T1. This led to a final sample for the ANT of 76 children (26 BilSs, 24 L2 learners and 26 NoL2s). Note that all other tasks were analysed on the full dataset of 99 participants. Before analysing Response Times (RT), we removed RTs faster than 200 ms (anticipatory responses) and slower than 2.5 standard deviations (SD) above the mean of each participant, which amounted to 19.5% of trials being excluded. RTs of incorrect responses were not considered in the analyses. Response omissions were recorded as errors.

Figure 3 shows the results for the conflict, alerting and orienting indices of the ANT task ($n = 76$). For the conflict index, there was no significant effect of Time, $F(1, 73) = .1$, $p = .739$, $\eta_p^2 = .002$, 95% CI [.00, .06], or Group, $F(2, 73) = 2.1$, $p = .131$, $\eta_p^2 = .054$, 95% CI [.00, .17], or a Time × Group interaction, $F(2, 73) = 1.2$, $p = .317$, $\eta_p^2 = .031$, 95% CI [.00, .13], meaning that there was no evidence for a significant improvement between T1 and T2 or differences between the groups.

For the alerting index, there was no effect of Time, $F(1, 73) < .01$, $p = .981$, $\eta_p^2 < .001$, 95% CI [.00, .00], or a Time × Group interaction, $F(2, 73) = 1.8$, $p = .117$, $\eta_p^2 = .046$, 95% CI [.00, .16], meaning that there was again no evidence for a significant improvement between T1 and T2 and no evidence for different developments across the groups. However, we found a significant main effect of Group, $n = 76$, $F(2, 73) = 3.9$, $p = .026$, $\eta_p^2 = .096$, 95% CI [.00, .24]. Bonferroni post-hoc tests showed that BilS children scored overall higher on the alerting index than L2 children, $t(72) = 2.69$, $p = .027$, Cohen's $d = 0.78$, 95% CI [0.18, 1.33], but not NoL2 children, $t(72) = 1.77$, $p = .242$, Cohen's $d = 0.51$, 95% CI [−.06, 1.05]. L2

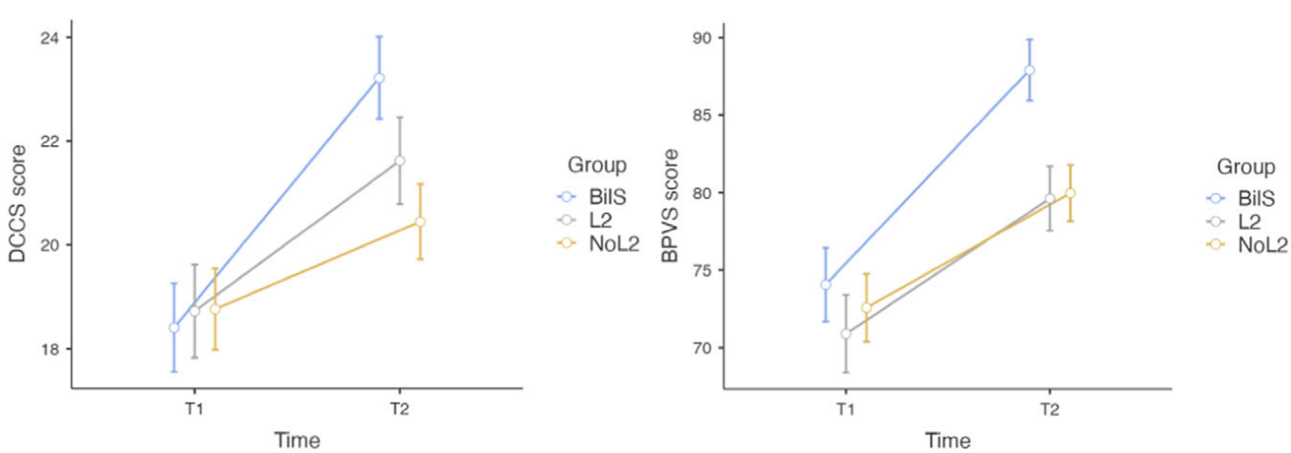

Figure 2. DCCS scores (left panel) and BPVS scores (right panel) for the three participant groups (BilS = bilingual school children, L2 = L2 learners, NoL2 = children without L2 provision) at both testing points (T1 and T2). Error bars represent standard errors.

https://doi.org/10.1017/S1366728925000069 Published online by Cambridge University Press

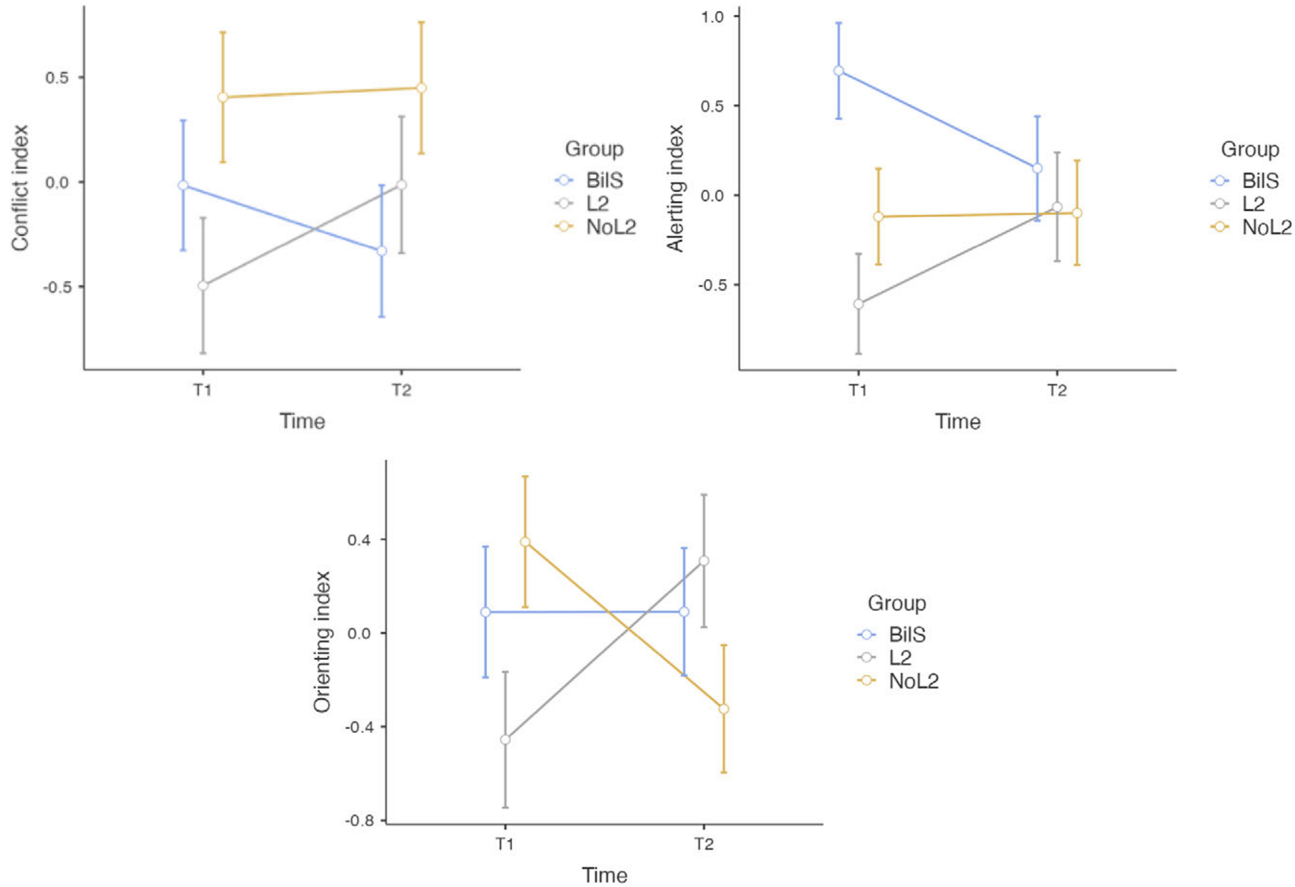

Figure 3. Mean conflict index (upper left panel), alerting index (upper right panel) and orienting index (lower middle panel) for the three participant groups (BilS = bilingual school children, L2 = L2 learners, NoL2 = children without L2 provision) at both testing points (T1 and T2). Error bars represent standard errors.

children did not differ from NoL2 children, $t(72) = .93$, $p = 1.00$, Cohen's $d = -.27$, 95% CI [−.83, .30].

For the orienting index, we found no effect of Time, $F(1, 73) < 0.1$, $p = .938$, $\eta_p^2 < .001$, 95% CI [.00, .00], or of Group, $F(2, 73) = .17$, $p = .847$, $\eta_p^2 = .005$, 95% CI [.00, .05], but a significant Time × Group interaction, $F(2, 73) = 3.61$, $p = .032$, $\eta_p^2 = .090$, 95% CI [.00, .23]. Figure 3 suggests that this interaction is due to different trends over time for the three groups. However, groups did not differ at T1, $F(2, 75) = 2.23$, $p = .119$, $\eta_p^2 = .06$, 95% CI [.00, .17], or at T2, Welch $F(2, 47.3) = 1.13$, $p = .331$, $\eta_p^2 = .05$, 95% CI [.00, .19]. Also, groups generally did not show any significant differences between T1 and T2 with Bonferroni-corrected $\alpha = 0.017$, BilS: $t(25) = 0.006$, $p = .990$, Cohen's $d < 0.01$, 95% CI [−.38,.39], L2: $t(23) = 2.13$, $p = .044$, Cohen's $d = 0.43$, 95% CI [.01,.85], NoL2: $t(25) = -1.38$, $p = .179$, Cohen's $d = -.27$, 95% CI [−.66,.12].

3.3. British Picture Vocabulary Scale (BPVS)

The right panel of Figure 2 shows the results of the BPVS ($n = 99$). We found a main effect of Time, $F(1, 96) = 103.9$, $p < .001$, $\eta_p^2 = .520$, 95% CI [.37, .64], no effect of Group, $F(2, 96) = 2.3$, $p = .108$, $\eta_p^2 = .045$, 95% CI [.00, .14], but a Time × Group interaction, $F(2,96) = 4.1$, $p = .019$, $\eta_p^2 = .079$, 95% CI [.00, .19]. All groups showed a significant improvement between T1 and T2, BilS: $t(31) = 7.2$, $p < .001$, Cohen's $d = 1.27$, 95% CI [0.80, 1.74], L2: $t(28) = 5.3$, $p < .001$, Cohen's $d = .98$, 95% CI [.53, 1.42], NoL2: $t(37) = 5.0$, $p < .001$, Cohen's $d = .81$, 95% CI [.44, 1.17]. However, while groups did not differ at T1, $F(2, 96) = .42$, $p = .657$, $\eta_p^2 = .01$, 95% CI [.00, .06], they did differ at T2, Welch $F(2, 60.9) = 6.82$, $p = .002$, $\eta_p^2 = .18$, 95% CI [.03, .36]. Games-Howell post-hoc tests for T2 showed higher BPVS scores for BilS children than L2 learners, $t(50.1) = 2.90$, $p = .015$, Cohen's $d = .76$, 95% CI [.22,1.26], and NoL2 children, $t(67.7) = 3.22$, $p = .006$, Cohen's $d = .78$, 95% CI [.28, 1.26], while L2 learners and NoL2 children did not differ, $t(57.2) = -.117$, $p = 1.0$, Cohen's $d = -.03$, 95% CI [−.51, .45]. Thus, BilS children improved more strongly than L2 learners and NoL2 children on the BPVS.

3.4. Director task

Split-half reliability tests for the director task at T1 and T2 showed good to excellent reliability (Cronbach's alpha .742 and .796, respectively). In addition, despite a large lag between T1 and T2, test–retest reliability across the two testing points was good, particularly for children without a language provision, Pearson $r(36) = .744$, $p < .001$, 95% CI [.55, .86], and children in bilingual schools, Pearson $r(30) = .693$, $p < .001$, 95% CI [.45, .84], while it was somewhat lower for weekly language learners, Pearson $r(27) = .538$, $p = .003$, 95% CI [.21, .76].

Figure 4 shows the results of the director task ($n = 99$). We found no effect of Group, $F(2, 96) = 2.1$, $p = .126$, $\eta_p^2 = .042$, 95% CI [.00, .14], but a main effect of Time, $F(1, 96) = 67.2$, $p < .001$, $\eta_p^2 = .412$, 95% CI [.26, .55], and a Time x Group interaction, $F(2, 96) = 14.9$, $p < .001$; $\eta_p^2 = .237$, 95% CI [.09, .38]. Only BilS children and L2

https://doi.org/10.1017/S1366728925000069 Published online by Cambridge University Press

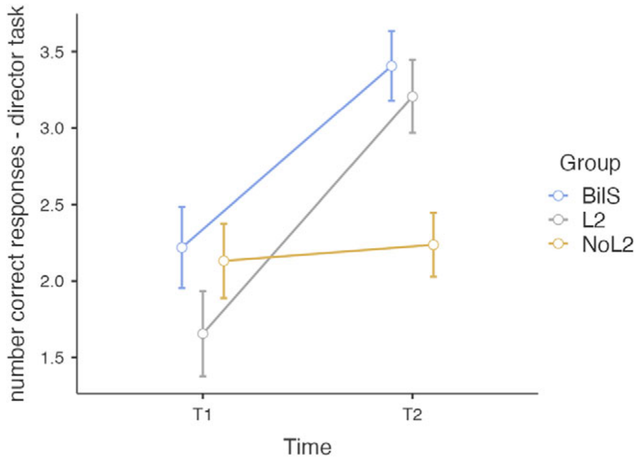

Figure 4. Mean number of correct responses to critical trials (max 4) in the director task for the three participant groups (BilS = bilingual school children, L2 = L2 learners, NoL2 = children without L2 provision) at both testing points (T1 and T2). Error bars represent standard errors.

learners showed a significant improvement between T1 and T2, BilS: $t(31) = 7.0$, $p < .001$, Cohen's $d = 1.24$, 95% CI [.77, 1.69], L2: $t(28) = 6.4$, $p < .001$, Cohen's $d = 1.19$, 95% CI [.70, 1.66], NoL2: $t(37) = .6$, $p = .578$, Cohen's $d = .10$, 95% CI [−.22, .42]. Also, there were no differences between groups at T1, $F(2, 96) = 1.3$, $p = .293$, $\eta_p^2 = .03$, 95% CI [.00, .11], but at T2, Welch F (2, 62.92) = 7.1, $p = .002$, $\eta_p^2 = .18$, 95% CI [.03, .35]. Games-Howell post-hoc tests for T2 showed higher perspective-taking scores for BilS children than NoL2 learners, $t(56.1) = .739$, $p = .001$, Cohen's $d = .91$, 95% CI [.40, 1.39], and L2 learners than NoL2 children, $t(64.5) = 2.92$, $p = .013$, Cohen's $d = .72$, 95% CI [.22, 1.22], while scores for BilS children and L2 learners did not differ, $t(56.1) = .739$, $p = .741$, Cohen's $d = .19$, 95% CI [−.31, .69].

3.5. Mediation analyses

BilS and L2 learners showed similar improvement on the Director task while NoL2 children showed no improvement. This pattern differs from those observed for the DCCS, BPVS and ANT, suggesting that children's development of communicative perspective-taking skills was not related to the development of their cognitive control skills or their language skills. To confirm this, we conducted mediation analyses. For that, we first established whether the improvement of performance (scores at T2 minus scores at T1) in the director task correlated with the performance difference across time points (scores at T2 minus scores at T1) in EF tasks and the BPVS. We combined the data of the three groups and removed the data from one participant in the NoL2 group who had a much lower improvement score (<3 SD) in the director task than any other participant.

The performance improvement in the director task did not correlate with the improvement in any of the ANT indices ($n = 75$), conflict index: Pearson $r(73) = -.03$, $p = .776$, 95% CI [−.26,.20], alerting index: Pearson $r(73) = .18$, $p = .128$, 95% CI [−.05,.39], orienting index: Pearson $r(73) = .19$, $p = .106$, 95% CI [−.04,.40]. Just as for the ANT, a partial correlation between performance improvement in the director task and the DCCS, controlling for age, showed no relationship either, Pearson $r(98) = .16$, $p = .116$, 95% CI [−.04,.35]. Finally, we found no significant correlation between performance improvement in the director task and performance improvement on the BPVS, Pearson $r(97) = .06$, $p = .585$, 95% CI [−.14,.25] (see confirming results for each participant group and each task in the Supplementary Material). Thus, correlation analyses confirmed that improvements on the director task were not related to improvements/changes on the EF tasks or on the BPVS. Since such relationships are necessary conditions for mediation to be present, we can conclude that improvement on the director task was not mediated by development in EF or language.

4. Discussion

We found that originally monolingual children who learned an L2 in an educational setting, either through attendance of a bilingual school or through weekly L2 lessons, showed improved socio-communicative perspective-taking skills compared to children without an L2 provision. Comparable improvements for children attending bilingual schools and those receiving weekly short lessons suggest that socio-communicative perspective-taking can be boosted effectively through moderate levels of L2 education.

4.1. Comparison to previous studies on precocious bilingual ToM skills

These results are congruent with the finding that children merely growing up in a multilingual environment from very early on without speaking the L2 themselves, thus at most passive bilinguals, performed comparably in the director task to children growing up as native bilinguals (Fan et al., 2015). They are also in line with Liberman et al. (2017) study which found evidence for a bilingual advantage among infants on a simplified version of the director task. However, our results stand somewhat in contrast with two cross-sectional studies that used false belief tasks and different age groups. Cheung et al. (2010) reported that full immersion into an L2 preschool led to higher ToM performance than five hours of L2 lessons per week, while Buac and Kauschkanskaya (2020) reported no difference in false-belief reasoning between monolingual children (~90 months) and children attending a bilingual ('dual-immersion') school. Since ToM abilities change rapidly in early childhood (e.g., Apperly, 2011; Wellman, 2014), it is possible that the amount of L2 exposure has different effects at different ages. Likewise, L2 exposure might affect different ToM tasks differently, either because they assess different facets of ToM or because they vary in their sensitivity for detecting differences. Note that the task employed here shows sensitivity to developmental differences across a wide age range (e.g., Dumontheil et al., 2010) compared with false belief tasks (e.g., Wellman et al., 2001), and so has the potential to detect L2 advantages that are long-lasting. Notably, ours is the first longitudinal study on the effects of L2 education on communicative perspective-taking (or ToM more generally), which ensured that groups were matched in their cognitive abilities before any L2 exposure.

4.2. The relationship of the development of communicative perspective-taking skills with EF and language development

Furthermore, we investigated whether performance development on the director task would be related to developments in EF or language. While there is ample evidence that individual differences in ToM are related to variations in both EF skills (Devine & Hughes, 2014) and language abilities (Milligan et al., 2007), previous research with bilingual children has provided somewhat mixed evidence for such relations (e.g., Buac & Kaushanskaya, 2020; Dahlgren et al., 2017; Diaz &

https://doi.org/10.1017/S1366728925000069 Published online by Cambridge University Press

Farrar, 2018b, 2018a; Fan et al., 2015; Kaushanskaya et al., 2014; Liberman et al., 2017; Nguyen & Astington, 2014). The present study showed that all groups improved in cognitive flexibility (see DCCS) and language development (see BPVS), with children in bilingual schools outperforming the other two groups at the second testing point and the scores of the latter not being significantly different. These results stand in contrast with the results of the director task, where children receiving weekly lessons patterned with children attending bilingual schools. The dissociation of the development in the director task versus EF tasks was confirmed by the finding that children's performance improvement in the director task did not correlate with their improvement in EF skills. The latter was the case both when taking all participant groups together and when examining the groups separately. The present findings, therefore, suggest that accelerated socio-communicative perspective-taking through L2 learning was independent of EF or language development and was unlikely caused by accelerated progression of EF or language abilities.

A similar dissociation between L2 language skills and communication perspective-taking skills is evident in recent research by Navarro et al. (2022) who investigated what aspects of bilingual experience affect performance on the director task in adults. They found that current experience and usage of L2 (i.e., frequency of usage and language switching) as well as language use of the participants during childhood (i.e., the usage of different languages in the family) predicted performance on the task. In contrast, performance was not predicted by current language fluency or metalinguistic awareness. Our results are only partly in line with these findings. On the one hand, given that children in the bilingual schools very likely had a higher L2 proficiency at T2 than weekly language learners, the results of similar performance in the two groups in the director task support the conclusion of Navarro et al. (2022) that L2 proficiency is not a strong predictor of ToM performance. On the other hand, though, the amount of exposure to a second language, as well as frequency of language usage, was not predictive of performance on the director task in our study. It might be that the latter is more important for adult performance than the enhancement of perspective-taking skills during early childhood. It might also be that the differences in measures in the two studies might have led to different results. Navarro and colleagues measured mouse trajectories in a computerised version of the director task, with a visual depiction of an avatar, while we measured accuracy in a real-life version of the task with a human communication partner. Our version therefore had a stronger social-communicative aspect. Furthermore, because Navarro et al. (2022) did not measure participants' cognitive control abilities, it is unclear whether their results might have been more strongly affected by mediating executive function differences than the present results.

One finding of our study might be rather surprising, namely that children attending bilingual schools had larger vocabularies towards the end of the first year at primary schools compared to the other two groups of children. Given that research with native bilingual children usually shows that children growing up with two languages simultaneously typically develop each of their two vocabularies more slowly (Bialystok & Senman, 2004; Carlson & Meltzoff, 2008; Chan, 2004; Diaz & Farrar, 2018a, 2018b; Farhadian et al., 2010; Nguyen & Astington, 2014; Nicoladis & Genesee, 1997), one might expect children attending bilingual primary schools will fall behind in their L1 development. However, other studies also suggest that this may not always be the case. For instance, Rhys and Thomas (2013) found that English monolingual children attending English-only schools had very similar vocabularies at the end of primary school as English–Welsh bilingual children who acquired Welsh through attendance at bilingual Welsh–English schools with dominant Welsh curriculum. Thus, bilingual education does not seem to have negative effects on L1, and English input, both outside and inside the classroom, seems sufficient to develop L1 vocabulary to the same level as attending a monolingual school. But why did children in bilingual schools accelerate their L1 vocabulary development faster compared to weekly language learners and monolingual controls? We can only speculate why this might be the case. It is unlikely that the bilingual schools focused more strongly on L1 vocabulary teaching, especially given that they had only half the time to do so compared to the other two school types. Instead, it might be that the L2 acquisition positively affected L1 acquisition, for instance, mediated through the development of metalinguistic skills. This is in line with evidence that higher metalinguistic skills are related to larger vocabularies (Altman et al., 2018). It also fits with Cummins' (1979) suggestion that L1 and L2 influence each other's development. Thus, it might be that, at least in the first year of bilingual education at primary school, education in L2 boosts L1 acquisition. Future research will need to replicate this finding.

4.3. What caused the accelerated development of communicative perspective-taking skills?

If the boost of communicative perspective-taking skills by L2 learning is not driven by a boost in EF development, the question arises of what causes the acceleration. It has been argued that sociolinguistic experiences drive the development of ToM (Díaz, 2022). The experience of communicating with an interlocutor who speaks another language, in other words, somebody who uses a different language to express their thoughts, feelings and desires, trains children's awareness that other people have different mental representations and knowledge. Learning to communicate with such a person means encountering communication difficulties and learning to pay attention to the knowledge and perspective of interlocutors. Native bilingual children learn to pay attention to the language knowledge of communication partners from early on. Bilingual children as young as two years pick the language appropriate to others' language knowledge (Genesee et al., 1996; Genesee et al., 1995; Nicoladis & Genesee, 1996) and are sensitive to communication cues from the communicator as to whether to mix languages or not (Lanza, 1992). Related to the language context in the present study, foreign language learners will use their common first language (English in the present study) when communicating with a friend at school. However, when communicating with a teacher who speaks another language, children need to consider that the teacher expects them to speak in 'their' language. Furthermore, foreign language lessons teach how a person from another country uses a different language to express their thoughts, feelings, and desires, as well as how they differ in their customs and ideas. Every lesson therefore stresses that other people might not have the same knowledge or experiences that the children have themselves. These experiences can enable young bilingual children and children learning another language in an educational setting to build a precocious understanding of the mind. This helps them to more successfully take into account another person's perspective. The latter is in line with our finding of an advanced performance in the director task. And it can potentially explain why bilingual children have been found to pass false-belief tasks at an earlier age than monolingual children (e.g., Berguno & Bowler, 2004; Diaz & Farrar, 2018a; Nguyen & Astington, 2014).

https://doi.org/10.1017/S1366728925000069 Published online by Cambridge University Press

Native bilingual children, i.e., those growing up with two languages at home, are likely not only exposed to different languages but also to different cultures and traditions. It might therefore not only be the experience with speakers of different languages that trains children's attentiveness to different perspectives, but also experiences with different cultures and traditions. While we do not know the exact details of the curriculum of any of the schools that took part in our study, it is likely that children with language provisions were not only exposed to a second language but also to cultural aspects, such as information about traditions and festivals related to the other language. This is a common practice in L2 language teaching in the United Kingdom, and especially bilingual schools focus on the L2 culture. Future studies will need to investigate how far the sociolinguistic aspect or the cultural exposure of the L2 lessons affected the development of communicative perspective-taking skills, and ToM development more generally.

4.4. Can we generalise the findings to other subtypes of ToM and other types of L2 education?

It has been argued that ToM is not monolithic, but a multidimensional construct (Navarro, 2022; Warnell & Redcay, 2019). Evidence for this comes from the finding that different measures of ToM do not necessarily correlate with each other (Warnell & Redcay, 2019). Our findings raise the question of whether language learning in educational settings leads to the same boost in performance on other ToM tasks. Buac and Kauschkanskaya (2020) did not find an advantage of children attending a bilingual school over monolingual children when testing them on a false-belief task. While this might suggest that false-belief reasoning might not be affected by language learning in an educational setting, their finding contrasts with that of Cheung et al. (2010) who reported that the level of L2 exposure and immersion in educational settings affects false belief reasoning, as they found that full immersion into an L2 preschool led to higher performance on false belief reasoning tasks than five hours of L2 lessons per week. It is therefore too early to conclude that L2 learning in educational settings might only affect some types of ToM abilities.

Apart from the question of whether ToM is multidimensional or monolithic, common ToM tasks do not only measure ToM but have components that are sensitive to general cognitive abilities, such as inhibition, memory and language (Coyle et al., 2018; Navarro, 2022; Warnell & Redcay, 2019). On the other hand, ToM tasks do not merely measure cognitive abilities or intelligence either. ToM has therefore been argued to be best characterised as a multidomain construct (Navarro, 2022). Many tasks measure lower-level cognitive processes. Even though Quesque and Rossetti (2020) argued that, compared to other tasks, the director task allows for an efficient exclusion of low-level interpretations of participants' performance (e.g., motor contagion), the director task has been argued to measure mental rotation or selective attention (Rubio-Fernández, 2017; Santiesteban et al., 2015). Navarro (2022) found that adults' director task performance correlated, even though only weakly, with performance on other ToM tasks such as the Reading the Mind in the Eyes task and a Short Stories Questionnaire that assesses inference of mental states of others. In addition, the director task also correlated with fluid intelligence. The latter suggests that the director task does not merely capture ToM. However, relationships with other ToM tasks in Navarro (2022) and the representation of the directors' view in Rubio-Fernándes (2017) suggest that it is more than a measure of general cognitive abilities. This conclusion is also supported by our finding that performance development in the director task was not related to the development of general cognitive ability. It should also be mentioned that studies such as Navarro (2022) and Santiesteban et al. (2015) used a computerised version of the director task with a visual depicture of an avatar. In contrast, in our task, participants interacted with a real person in a real social interaction. This might encourage mentalising more strongly.

It is important to consider whether our findings can be generalised to any L2 learning in educational settings. In other words, might differences in communicative perspective-taking skills between our groups have arisen due to other group differences than differences in L2 teaching exposure? Crucially, the groups were matched in terms of personal measures (e.g., family background or extra-curricular activities) and initial cognitive abilities.[2] This tight control of confounding variables suggests that differences in experiences that children had outside the school unlikely affected their performance at the second testing point. The more likely source of differences is experiences at school. Since all schools followed the national curriculum, our results could in principle, be due to differences in teaching implementation or teaching effectiveness by the different types of schools. But two reasons speak against this. First, children in each group were recruited from at least two schools, meaning that implementation of the national curriculum will have varied among schools of each type. Second, if we allow that English vocabulary development of the children's native language can serve as an indication of teaching effectiveness, it is notable that it showed a different pattern of results than that for the director task. Bilingual school children showed accelerated vocabulary development compared to the other two groups. Thus, if anything, only the teaching methods of bilingual schools, but not of schools with weekly L2 provisions, might have been more effective. Thus, it is unlikely that the acceleration of perspective-taking development in *both* groups was due to higher teaching effectiveness. Having said this, we cannot completely rule out that the school types systematically differed in aspects that would have affected communicative perspective-taking development. Future research should therefore try to replicate our findings in a randomised controlled study that tightly controls for potential confounding variables. In addition, it would be useful to complement the director task with other ToM tasks such as false belief tasks to investigate whether learning another language might affect performance in a ToM task with social-communicative aspects more strongly than ToM tasks that lack such aspect.

4.5. Limitations

Our study has several limitations. Its main limitation is its non-randomised design. Despite the groups being very similar in family backgrounds, not all aspects of family background and school characteristics might have been controlled for. For instance, one aspect that we did not measure was the amount of parental involvement in children's education, which might have led to the decision to send children to a bilingual school. One caveat here is that bilingual schools are very rare in the United Kingdom and that weekly language learners and control children did not have access to a bilingual school. Nevertheless, there might

[2]An exception was that children without a foreign language provision were older than the other two groups, especially bilingual school children. However, age did not correlate with performance in the director task and therefore cannot explain the results.

https://doi.org/10.1017/S1366728925000069 Published online by Cambridge University Press

be differences between the groups that were not controlled for. Similarly, we did not gather detailed information about the teaching methods of the schools. For instance, we do not know how far any of the schools exposed the children to different cultures. It would therefore be important to record such details in future studies and to replicate the current findings in a randomised controlled study.

Another limitation of our study is that we did not test children's L2 ability at the second testing point. While the finding that children in bilingual schools and weekly language learners improved in their perspective-taking skills to the same degree, it would still be interesting to see how far the level of L2 proficiency might correlate with perspective-taking development. Also, given the strong increase in L1 vocabulary in bilingual school children and our suggestion that this might be due to an increase in metalinguistic skills, it would also have been useful to know whether metalinguistic skills would indeed have increased in this group more strongly than in the two other participant groups.

Furthermore, we tested the children only on a single measure of ToM because we were interested in their social communicative abilities. We are not aware of another study that compared the performance of bilingual or monolingual children on the director task with performance on other common ToM tasks such as false belief tasks. It would be interesting to see whether performance in these tasks correlates, once language or EF skills are controlled for. If so, L2 learning might lead to accelerations of ToM development more generally. And it would provide evidence against the concern that the bilingual advantage in ToM might be rather an advantage in attentional control or cognitive skills in general (Rubio-Fernandez, 2016, 2017). Such a study would also be able to test whether the social aspect of the director task might lead to larger effects of L2 learning than other ToM tasks.

4.6. Conclusion

The results of the present study have potential societal implications, especially if we assume that our findings for the director task are representative of ToM development more generally instead of a particular type of ToM. As indicated, ToM is a foundation for social competence. For instance, ToM of six-year-olds predicts teacher ratings of social skills at age 10 (Devine et al., 2016). In relation to this, it is interesting that children from a minority language background (thus native bilinguals) are rated higher by teachers and observers on measures of interpersonal skills, compared with monolingual children who speak the majority language (Halle et al., 2014). This might be due to their superior ToM skills. If language learning in an educational setting indeed fosters social competence, then language learning could have long-lasting effects on children's lives. For example, for school-aged children, good social competence predicts future mental health, academic outcomes and success in the workplace (Bornstein et al., 2010). On a societal level, language learning might even lead to a society with less xenophobia and aggression against minorities. It has been found that individuals with superior perspective-taking skills more easily suppress automatic expressions of racial bias (Todd et al., 2011). Developing social skills through L2 learning is therefore a potential tool to cultivate a civic identity in children and to build a functioning diverse society. Furthermore, L2 learning at school in our study was effective even at a low dose, meaning that benefits might be achievable without major modifications to the standard school curriculum. However, future studies will need to explore any long-term effects of early second language provisions.

To conclude, our study that children who learned a second language in early childhood in a formal educational setting showed enhanced communicative perspective-taking skills. A high amount of L2 exposure does not seem to be necessary for such a boost to occur because short weekly lessons led to the same acceleration as attendance at a bilingual school. Our results present an argument for early language teaching in schools.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S1366728925000069.

Data availability statement. All data, analysis and materials for the experiment are available at the open science framework at https://osf.io/s9ghz/?view_only=d3d0bdf8a5a443a581d78b4ff51e746b

Acknowledgement. We would like to thank the part-taking schools, children and their parents.

Competing interest. The authors declare no conflicts of interest.

Ethics approval statement. The study protocol received full ethical approval from the Science, Technology, Engineering and Mathematics Ethics Committee of the University of Birmingham, United Kingdom (ERN_17–0324).

References

Adesope, O., **Lavin, T.**, **Thompson, T.**, & **Ungerleider, C.** (2010). *A systematic review and meta-analysis of the cognitive correlates of bilingualism* (Vol. **80**). https://doi.org/10.3102/0034654310368803

Altman, C., **Goldstein, T.**, & **Armon-Lotem, S.** (2018). Vocabulary, metalinguistic awareness and language dominance among bilingual preschool children. *Frontiers in Psychology*, **9**. https://doi.org/10.3389/fpsyg.2018.01953

Apperly, I. A. (2011). *Mindreaders: The cognitive basis of "theory of mind"*. Psychology Press.

Beck, D. M., **Schaefer, C.**, **Pang, K.**, & **Carlson, S. M.** (2011). Executive function in preschool children: Test–retest reliability. *Journal of Cognition and Development*, **12**(2), 169–193. https://doi.org/10.1080/15248372.2011.563485

Berguno, G., & **Bowler, D. M.** (2004). Communicative interactions, knowledge of a second language, and theory of mind in young children. *The Journal of Genetic Psychology*, **165**(3), 293–309. https://doi.org/10.3200/GNTP.165.3.293-309

Bialystok, E., & **Senman, L.** (2004). Executive processes in appearance-reality tasks: The role of inhibition of attention and symbolic representation. *Child Development*, **75**(2), 562–579. https://doi.org/10.1111/j.1467-8624.2004.00693.x

Bornstein, M. H., **Hahn, C.-S.**, & **Haynes, O. M.** (2010). Social competence, externalizing, and internalizing behavioral adjustment from early childhood through early adolescence: Developmental cascades. *Development and Psychopathology*, **22**(4), 717–735. https://doi.org/10.1017/S0954579410000416

Boyd, S., & **Ottesjö, C.** (2016). Adult monolingual policy becomes children's bilingual practice: Code-alternation among children and staff in an English-medium preschool in Sweden. *International Journal of Bilingual Education and Bilingualism*, **19**(6), 631–648. https://doi.org/10.1080/13670050.2016.1189510

Buac, M., & **Kaushanskaya, M.** (2020). Predictors of theory of mind performance in bilingual and monolingual children. *International Journal of Bilingualism*, **24**(2), 339–359. https://doi.org/10.1177/1367006919826866

Carlson, S. M., & **Meltzoff, A. N.** (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, **11**(2), 282–298. https://doi.org/10.1111/j.1467-7687.2008.00675.x

Casagrande, M., **Marotta, A.**, **Martella, D.**, **Volpari, E.**, **Agostini, F.**, **Favieri, F.**, **Forte, G.**, **Rea, M.**, **Ferri, R.**, **Giordano, V.**, **Doricchi, F.**, & **Giovannoli, J.** (2022). Assessing the three attentional networks in children from three to six years: A child-friendly version of the Attentional Network Test for Interaction. *Behavior Research Methods*, **54**(3), 1403–1415. https://doi.org/10.3758/s13428-021-01668-5

Chan, K. T. (2004). *Chine-English Bilingual's theory of mind development.* (PhD thesis). Toronto University, Toronto.

Cheung, H., **Mak, W. Y.**, **Luo, X.**, & **Xiao, W.** (2010). Sociolinguistic awareness and false belief in young Cantonese learners of English. *Journal of*

Experimental Child Psychology, **107**(2), 188–194. https://doi.org/10.1016/j.jecp.2010.05.001

Christoffels, I. K., **de Haan, A. M.**, **Steenbergen, L.**, **van den Wildenberg, W. P. M.**, & **Colzato, L. S.** (2015). Two is better than one: bilingual education promotes the flexible mind. *Psychological Research*, **79**(3), 371–379. https://doi.org/10.1007/s00426-014-0575-3

Coyle, T. R., **Elpers, K. E.**, **Gonzalez, M. C.**, **Freeman, J.**, & **Baggio, J. A.** (2018). General intelligence (g), ACT scores, and theory of mind: (ACT)g Predicts limited variance among theory of mind tests. *Intelligence*, **71**, 85–91. https://doi.org/10.1016/j.intell.2018.10.006

Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, **19**, 121–129. https://coilink.org/20.500.12592/667ug1k

Dahlgren, S. O., **Almén, H.**, & **Dahlgren Sandberg, A.** (2017). Theory of mind and executive functions in young bilingual children. *Journal of Genetic Psychology*, **178**(5), 303–307. https://doi.org/10.1080/00221325.2017.1361376

DeLuca, V., **Segaert, K.**, **Mazaheri, A.**, & **Krott, A.** (2020). Understanding bilingual brain function and structure changes? U bet! A unified bilingual experience trajectory model. *Journal of Neurolinguistics*, **56**, 100930. https://doi.org/10.1016/j.jneuroling.2020.100930

DeLuca, V., **Voits, T.**, **Ni, J.**, **Carter, F.**, **Rahman, F.**, **Mazaheri, A.**, **Krott, A.**, & **Segaert, K.** (2024). Mapping individual aspects of bilingual experience to adaptations in brain structure. *Cerebral Cortex*, **34**(2). https://doi.org/10.1093/cercor/bhae029

Devine, R. T., & **Hughes, C.** (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, **85**(5), 1777–1794. https://doi.org/10.1111/cdev.l2237

Devine, R. T., **White, N.**, **Ensor, R.**, & **Hughes, C.** (2016). Theory of mind in middle childhood: Longitudinal associations with executive function and social competence. *Developmental Psychology*, **52**(5), 758–771. https://doi.org/10.1037/dev0000105

Díaz, V. (2022). Minds in action: Evidence that linguistic diversity helps children build a theory of mind. *Bilingualism: Language and Cognition*, **25**(1), 70–80. https://doi.org/10.1017/S1366728921000109

Diaz, V., & **Farrar, M. J.** (2018a). Do bilingual and monolingual preschoolers acquire false belief understanding similarly? The role of executive functioning and language. *First Language*, **38**(4), 382–398. https://doi.org/10.1177/0142723717752741

Diaz, V., & **Farrar, M. J.** (2018b). The missing explanation of the false-belief advantage in bilingual children: A longitudinal study. *Developmental Science*, **21**(4), 1–13. https://doi.org/10.1111/desc.12594

Donnelly, S., **Brooks, P. J.**, & **Homer, B. D.** (2019). Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost. *Psychonomic Bulletin & Review*, **26**(4), 1122–1147. https://doi.org/10.3758/s13423-019-01567-z

Dumontheil, I., **Apperly, I. A.**, & **Blakemore, S. J.** (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, **13**(2), 331–338. https://doi.org/10.1111/j.1467-7687.2009.00888.x

Dunn L. M., **Dunn L. M.**, **Whetton C.** & **Burley J.** (1997). *The British Picture Vocabulary Scale*, 2nd edn. National Foundation for Educational Research. NFER-NELSON, Berkshire.

Dunn, L. M., **Dunn, D. M.**, **Sewell, J.**, **Styles, B.**, **Brzyska, B.**, **Shamsan, Y.**, & **Burge, B.** (2009). *British Picture Vocabulary Scale* (3rd ed.). GL Assessment Limited.

Elvin, P., **Maagerø, E.**, & **Simonsen, B.** (2007). How do the dinosaurs speak in England? English in kindergarten. *European Early Childhood Education Research Journal*, **15**(1), 71–86. https://doi.org/10.1080/13502930601103199

Fan, S. P., **Liberman, Z.**, **Keysar, B.**, & **Kinzler, K. D.** (2015). The exposure advantage: Early exposure to a multilingual environment promotes effective communication. *Psychological Science*, **26**(7), 1090–1097. https://doi.org/10.1177/0956797615574699

Farhadian, M., **Abdullah, R.**, **Mansor, M.**, **Redzuan, M.**, **Gazanizadand, N.**, & **Kumar, V.** (2010). Theory of Mind in Bilingual and Monolingual Preschool Children. *Journal of Psychology*, **1**. https://doi.org/10.1080/09764224.2010.11885444

Gampe, A., **Wermelinger, S.**, & **Daum, M. M.** (2019). Bilingual children adapt to the needs of their communication partners, monolinguals do not. *Child Development*, **90**(1), 98–107. https://doi.org/10.1111/cdev.13190

Genesee, F., **Boivin, I.**, & **Nicoladis, E.** (1996). Talking with strangers: A study of bilingual children's communicative competence. *Applied Psycholinguistics*, **17**(4), 427–442. https://doi.org/10.1017/S0142716400008183

Genesee, F., **Nicoladis, E.**, & **Paradis, J.** (1995). Language differentiation in early bilingual development. *Journal of Child Language*, **22**(3), 611–631. https://doi.org/10.1017/S0305000900009971

Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition*, **6**(1), 1–15. https://doi.org/10.1017/S1366728903001007

Gordon, K. R. (2016). High proficiency across two languages is related to better mental state reasoning for bilingual children. *Journal of Child Language*, **43**(2), 407–424. https://doi.org/10.1017/s0305000915000276

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, **1**(2), 67–81. https://doi.org/10.1017/S1366728998000133

Green, D. W., & **Abutalebi, J.** (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, **25**(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Grundy, J. G. (2020). The effects of bilingualism on executive functions: an updated quantitative analysis. *Journal of Cultural Cognitive Science*, **4**(2), 177–199. https://doi.org/10.1007/s41809-020-00062-5

Halle, T. G., **Whittaker, J. V.**, **Zepeda, M.**, **Rothenberg, L.**, **Anderson, R.**, **Daneri, P.**, **Wessel, J.**, & **Buysse, V.** (2014). The social-emotional development of dual language learners: Looking back at existing research and moving forward with purpose. *Early Childhood Research Quarterly*, **29**(4), 734–749. https://doi.org/10.1016/j.ecresq.2013.12.002

Janus, M., **Lee, Y.**, **Moreno, S.**, & **Bialystok, E.** (2016). Effects of short-term music and second-language training on executive control. *Journal of Experimental Child Psychology*, **144**, 84–97. https://doi.org/10.1016/j.jecp.2015.11.009

Kalia, V., **Daneri, M. P.**, & **Wilbourn, M. P.** (2019). Relations between vocabulary and executive functions in Spanish-english dual language learners. *Bilingualism*, **22**(1), 1–14. https://doi.org/10.1017/S1366728917000463

Kaushanskaya, M., **Gross, M.**, & **Buac, M.** (2014). Effects of classroom bilingualism on task-shifting, verbal memory, and word learning in children. *Developmental Science*, **17**(4), 564–583. https://doi.org/10.1111/desc.12142

Keysar, B., **Barr, D. J.**, **Balin, J. A.**, & **Brauner, J. S.** (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, **11**(1), 32–38. https://doi.org/10.1111/1467-9280.00211

Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, **12**(1), 48–54. https://doi.org/10.1111/j.1467-7687.2008.00742.x

Lakens, D., & **Caldwell, A. R.** (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science*, **4**(1), 2515245920951503. https://doi.org/10.1177/2515245920951503

Lanza, E. (1992). Can bilingual two-year-olds code-switch? *Journal of Child Language*, **19**(3), 633–658. https://doi.org/10.1017/S0305000900011600

Lehtonen, M., **Soveri, A.**, **Laine, A.**, **Jarvenpaa, J.**, **de Bruin, A.**, & **Antfolk, J.** (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, **144**(4), 394–425. https://doi.org/10.1037/bul0000142

Liesefeld, H. R., **Fu, X.**, & **Zimmer, H. D.** (2015). Fast and careless or careful and slow? Apparent holistic processing in mental rotation is explained by speed-accuracy trade-offs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **41**(4), 1140–1151. https://doi.org/10.1037/xlm0000081

Liberman, Z., **Woodward, A. L.**, **Keysar, B.**, & **Kinzler, K. D.** (2017). Exposure to multiple languages enhances communication skills in infancy. *Developmental Science*, **20**(1). https://doi.org/10.1111/desc.12420

Lightbown, P. M., & **Spada, N.** (2013). *How languages are learned* (4th ed.). Oxford University Press.

Lowe, C. J., **Cho, I.**, **Goldsmith, S. F.**, & **Morton, J. B.** (2021). The bilingual advantage in children's executive functioning is not related to language status: A meta-analytic review. *Psychological Science*, **32**(7), 1115–1146. https://doi.org/10.1177/0956797621993108

Lucas, C., **Hood, P.**, & **Coyle, D.** (2021). Blossoming in English: Preschool children's emergent literacy skills in English. *Journal of Research in Childhood Education*, **35**(3), 477–502. https://doi.org/10.1080/02568543.2020.1742256

Luk, G., & **Bialystok, E.** (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, **25**(5), 605–621. https://doi.org/10.1080/20445911.2013.795574

Luna, F. G., **Roca, J.**, **Martín-Arévalo, E.**, & **Lupiáñez, J.** (2021). Measuring attention and vigilance in the laboratory vs. online: The split-half reliability of the ANTI-Vea. *Behavior Research Methods*, **53**(3), 1124–1147. https://doi.org/10.3758/s13428-020-01483-4

Milligan, K., **Astington, J. W.**, & **Dack, L. A.** (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, **78**(2), 622–646. http://www.jstor.org/stable/4139249

Murphy, V. A., **Arndt, H.**, **Briggs Baffoe-Djan, J.**, **Chalmers, H.**, **Macaro, E.**, **Heath, R.**, **Vanderplank, R.**, & **Woore, R.** (2020). *Foreign language learning and its impact on wider academic outcomes: A rapid evidence assessment.* https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/foreign-language-learning.

Navarro, E. (2022). What is theory of mind? A psychometric study of theory of mind and intelligence. *Cogn Psychol*, **136**, 101495. https://doi.org/10.1016/j.cogpsych.2022.101495

Navarro, E., **DeLuca, V.**, & **Rossi, E.** (2022). It takes a village: Using network science to identify the effect of individual differences in bilingual experience for theory of mind. *Brain Sciences*, **12**(4). https://doi.org/10.3390/brainsci12040487

Neveu, A., **Crespo, K.**, **Ellis Weismer, S.**, & **Kaushanskaya, M.** (2021). Does long-term dual-language immersion affect children's executive functioning? *Journal of Experimental Child Psychology*, **208**, 105127. https://doi.org/10.1016/j.jecp.2021.105127

Nguyen, T. K., & **Astington, J. W.** (2014). Reassessing the bilingual advantage in theory of mind and its cognitive underpinnings. *Bilingualism*, **17**(2), 396–409. https://doi.org/10.1017/S1366728913000394

Nicoladis, E., & **Genesee, F.** (1996). A longitudinal study of pragmatic differentiation in young bilingual children. *Language Learning*, **46**(3), 439–464. https://doi.org/10.1111/j.1467-1770.1996.tb01243.x

Nicoladis, E., & **Genesee, F.** (1997). Language development in preschool bilingual children. *Journal of Speech-Language Pathology and Audiology*, **21**, 258–270.

Nicolay, A.-C., & **Poncelet, M.** (2013). Cognitive advantage in children enrolled in a second-language immersion elementary school program for three years. *Bilingualism-Language and Cognition*, **16**(3), 597–607. https://doi.org/10.1017/s1366728912000375

Office for National Statistics. (2021). *2021 census profile for areas in England and Wales.* https://www.nomisweb.co.uk/sources/census_2021/report

Peristeri, E., **Baldimtsi, E.**, **Durrleman, S.**, & **Tsimpli, I.** (2019). Bilingualism effects in children with developmental language disorder: Metalinguistic awareness, executive functions and false belief reasoning. In M. M. Brown & B. Dailey (Eds.), *Proceedings of the 43rd Boston University conference on language development* (pp. 523–536). Boston: Cascadilla Press.

Pinter, A. (2017). *Teaching young language learners* (2nd ed.). Oxford University Press.

Pliatsikas, C. (2020). Understanding structural plasticity in the bilingual brain: The dynamic restructuring model. *Bilingualism: Language and Cognition*, 1–13. https://doi.org/10.1017/S1366728919000130

Quesque, F., & **Rossetti, Y.** (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science*, **15**(2), 384–396. https://doi.org/10.1177/1745691619896607

Rafeekh, R., **Krishna, P. P.**, **Kapiley, K.**, & **Mishra, R. K.** (2021). The effects of short-term L2 training on components of executive control in Indian bilinguals. *Cognitive Processing*, **22**(2), 339–351. https://doi.org/10.1007/s10339-021-01014-9

Raven, J. C., **Court, J. H.**, & **Raven, J.** (1998). *Coloured progressive matrices.* Oxford Psychologist Press.

Rhys, M., & **Thomas, E. M.** (2013). Bilingual Welsh–English children's acquisition of vocabulary and reading: implications for bilingual education. *International Journal of Bilingual Education and Bilingualism*, **16**(6), 633–656. https://doi.org/10.1080/13670050.2012.706248

Rubio-Fernandez, P. (2016). Why are bilinguals better than monolinguals at false-belief tasks? *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-016-1143-1

Rubio-Fernández, P. (2017). The director task: A test of Theory-of-Mind use or selective attention? *Psychonomic Bulletin & Review*, **24**(4), 1121–1128. https://doi.org/10.3758/s13423-016-1190-7

Rueda, M. R., **Fan, J.**, **McCandliss, B. D.**, **Halparin, J. D.**, **Gruber, D. B.**, **Lercari, L. P.**, & **Posner, M. I.** (2004). Development of attentional networks in childhood. *Neuropsychologia*, **42**(8), 1029–1040. https://doi.org/10.1016/j.neuropsychologia.2003.12.012

Santiesteban, I., **Shah, P.**, **White, S.**, **Bird, G.**, & **Heyes, C.** (2015). Mentalizing or submentalizing in a communication task? Evidence from autism and a camera control. *Psychonomic Bulletin & Review*, **22**(3), 844–849. https://doi.org/10.3758/s13423-014-0716-0

Schroeder, S. R. (2018). Do bilinguals have an advantage in theory of mind? A meta-analysis. *Frontiers in Communication*, **3**(August), 1–8. https://doi.org/10.3389/fcomm.2018.00036

Todd, A. R., **Bodenhausen, G. V.**, **Richeson, J. A.**, & **Galinsky, A. D.** (2011). Perspective taking combats automatic expressions of racial bias. *Journal of Personality and Social Psychology*, **100**(6), 1027–1042. https://doi.org/10.1037/a0022308

Warnell, K. R., & **Redcay, E.** (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition*, **191**, 103997. https://doi.org/10.1016/j.cognition.2019.06.009

Wellman, H. M., **Cross, D.**, & **Watson, J.** (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, **72**(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.

Yang, S., **Yang, H.**, & **Lust, B.** (2011). Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. *Bilingualism*, **14**(3), 412–422. https://doi.org/10.1017/S1366728910000611

Yelland, G. W., **Pollard, J.**, & **Mercuri, A.** (1993). The metalinguistic benefits of limited contact with a second language. *Applied Psycholinguistics*, **14**(4), 423–444. https://doi.org/10.1017/S0142716400010687

Yow, W. Q., & **Markman, E. M.** (2011). Bilingualism and children's use of paralinguistic cues to interpret emotion in speech. *Bilingualism: Language and Cognition*, **14**(4), 562–569. https://doi.org/10.1017/s1366728910000404

Yow, W. Q., & **Markman, E. M.** (2015). A bilingual advantage in how children integrate multiple cues to understand a speaker's referential intent. *Bilingualism: Language and Cognition*, **18**(3), 391–399. https://doi.org/10.1017/S1366728914000133

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, **1**(1), 297–301. https://doi.org/10.1038/nprot.2006.46

https://doi.org/10.1017/S1366728925000069 Published online by Cambridge University Press