The effect of language proficiency on theory of mind: A longitudinal study of bilingual autistic and neurotypical children

Yulia Zuban, Franziska Baumeister, Pauline Wolfer, Stephanie Durrleman

Child Development, 2026, 1–14
https://doi.org/10.1093/chidev/aacag092
Published: May 12, 2026
Empirical Article

The effect of language proficiency on theory of mind: A longitudinal study of bilingual autistic and neurotypical children

Yulia Zuban[1,2,*], Franziska Baumeister[1], Pauline Wolfer[1], and Stephanie Durrleman[1]

[1]ABCCD Lab, Faculty of Science and Medicine, University of Fribourg, Fribourg, Switzerland
[2]Department of English Linguistics, University of Stuttgart, Stuttgart, Germany
*Corresponding author: ABCCD Lab, Faculty of Science and Medicine, University of Fribourg, Fribourg, Switzerland, Chemin du Musée 5, Fribourg CH-1700, Switzerland; Department of English Linguistics, University of Stuttgart, Keplerstraße 17, Stuttgart 70174, Germany. Emails: yulia.zuban@unifr.ch, yulia.zuban@ling.uni-stuttgart.de

Abstract

The study examines how bilinguals' languages of higher proficiency (LHP) and lower proficiency (LLP) correlate with daily-life Theory of Mind (ToM) in 137 White, bilingual typically developing (TD; 45% boys) and 107 White, bilingual autistic children (78% boys) aged 6 to 12. A subset of participants (28 TD, 43% boys; 21 autistic children, 76% boys) was reassessed approximately 1 year later. Age positively correlated with ToM in TD and autistic children, although the former outperformed the latter. In autistic children, but not TD peers, LHP and LLP were both positively associated with greater ToM skills. Longitudinal findings 1 year later confirmed this pattern, suggesting that bilingual autistic children continue to capitalize on both languages over time to support ToM development.

Keywords autism, theory of mind, bilingualism, longitudinal study, language proficiency

Lay summary

This study explores how bilingual children use their stronger and weaker languages to understand emotions, desires, and beliefs in themselves and others (Theory of Mind). Overall, 137 White, bilingual typically developing (45% boys) and 107 White, bilingual autistic children (78% boys) aged 6 to 12 were evaluated, and 20% were reevaluated 1 year later. Theory of Mind improved with age for typically developing and autistic children, although the former outperformed the latter. Only in autistic children were stronger and weaker languages linked to better Theory of Mind. At follow-up a year later, the same pattern emerged, suggesting that bilingual autistic children continue to benefit from both languages over time, using them to support social and emotional growth.

Recent estimates suggest that over half of the global population speaks two or more languages (De Bruin, 2019). Bilingualism, broadly defined, refers to the use of multiple languages or dialects in daily life (Grosjean et al., 2013). The current study explores whether being exposed to more than one language correlates with Theory of Mind (ToM), the ability to understand and interpret mental states-including emotions, desires, and beliefs, in oneself and others (Baron-Cohen, 1995), in both bilingual typically developing children (further TD), and bilingual children diagnosed with autism spectrum disorder.

Language-related skills (i.e., lexical and syntactic abilities) have been identified as key predictors of ToM in TD children (De Villiers & Pyers, 2002; Farrar & Maag, 2002) and autistic children (Durrleman et al., 2016; Tager-Flusberg & Joseph, 2005). Further, although language proficiency in a stronger language has emerged as a robust predictor of improved performance on ToM tasks, only a few studies examined the relation between *two* languages and ToM abilities (see a review by Białecka et al., 2024). To date, the only longitudinal study exploring bilingualism and ToM is that of Peristeri et al. (2024), who investigated ToM using a behavioral, lab-based experiment rather than a daily-life task. This study is the first to investigate the possible link between proficiency in *two* languages of autistic and TD children and daily-life ToM over time.

Since the 1990s, research on bilingualism has grown significantly, exploring its potential impact on various cognitive

Handling Editor: Adam Winsler
Editor: Adam Winsler
Received: August 22, 2025. **Revised:** April 10, 2026. **Accepted:** April 13, 2026

domains, including executive functions (EFs) (Bialystok, 2011) and ToM (Schroeder, 2018), shedding light on the broader cognitive implications of bilingual experience. Executive functions encompass advanced cognitive control processes responsible for regulating goal-directed thinking, emotional responses, and behavioral actions (Diamond, 2013). While numerous studies have linked bilingualism to cognitive benefits in TD children (see a meta-analysis by Gunnerud et al., 2020), other studies have reported null results (Lowe et al., 2021).

The link between bilingualism and ToM is crucial for clinical populations, such as those diagnosed with autism spectrum disorder, since autistic children, being a heterogeneous group, are often reported to struggle with ToM (Yirmiya et al., 1998) or to differ in ToM from TD populations despite being able to represent mental states (Long et al., 2025). Autism spectrum disorder (ASD), one of the most widespread neurodevelopmental disorders today (Zeidan et al., 2022), typically affects social communication (American Psychiatric Association, 2022), potentially yielding delays and deficits in language development (Silleresi, 2023). Parents of autistic children are reportedly cautioned against bilingual upbringing, as it is believed that exposure to multiple languages may intensify certain challenges associated with ASD (Beauchamp & MacLeod, 2017; Hampton et al., 2017). Recent empirical research indicates that bilingual autistic children do not experience heightened language delays (Giovannoli et al., 2020). Additionally, avoiding bilingualism can negatively impact family bonding and overall well-being, issues that could be mitigated by maintaining bilingual exposure (Yu, 2013). Therefore, expanding the empirical understanding of bilingualism in this population is essential to support the autism community in making evidence-based decisions regarding multilingualism (Beauchamp & MacLeod, 2017).

Research on the association between bilingualism and ToM in TD children reports that bilingual children outperform their monolingual peers (Baumeister et al., 2025; Schroeder, 2018). Research on bilingualism and ToM in autistic children is limited. Some studies report on the bilingual advantage of autistic children in first-order (i.e., the ability to understand that others can hold incorrect beliefs about real-world situations) and second-order ToM tasks (i.e., the ability to understand that one person can hold a mistaken belief about another person's belief) (Andreou et al., 2020; Peristeri et al., 2024, 2021a, 2021b; Perner & Wimmer, 1985). However, there is also evidence that bilingual autistic children perform similarly to their monolingual peers on these tasks (Meir & Novogrodsky, 2019).

Reviews of existing behavioral ToM assessment tools (Beaudoin et al., 2020; Quesque & Rossetti, 2020) highlight significant challenges in reliably evaluating ToM in children. For instance, many assessments place heavy cognitive and verbal demands on children, making them particularly difficult for those with ASD. Additionally, existing ToM measures vary greatly in length, reliability, and feasibility: while shorter tasks often lack reliability, longer tasks present practical constraints. Finally, behavioral ToM tasks are limited in capturing ToM skills in everyday life, underscoring the need for more comprehensive and ecologically valid assessment approaches. As a supplementary assessment tool, parental questionnaires have been designed to evaluate a wider spectrum of ToM skills in everyday life. The current study is one of a few existing studies to examine the informant-based ToM questionnaire ToMI-2 (Hutchins et al., 2012), providing insights beyond behavioral experimental tasks. Importantly, the ToMI-2 demonstrates strong psychometric properties, including high internal consistency and excellent test-retest reliability (internal consistency, Cronbach's $\alpha = .98$; test-retest reliability, $M = 27.5$; $SD = 19.4$; $r = .89$, $p < .001$) (Hutchins et al., 2012, p. 331). The ToMI-2 also effectively distinguishes between children based on age and developmental status (ASD vs. TD; Baumeister et al., 2024). Scores on the ToMI-2 correlate with performance on traditional ToM task batteries, supporting its construct validity (Hutchins et al., 2012).

Recent theoretical work has underscored the limits of many existing ToM measures for drawing inferences about underlying cognitive mechanisms in autism (Long et al., 2025). In this context, the present study contributes to the growing, yet still limited, body of research on autism and bilingualism by investigating how the degree of proficiency (i.e., higher or lower) in *two* languages correlates with ToM in daily life in bilingual autistic and TD children. ToM in daily life is measured through the parental questionnaire ToMI-2, a validated assessment tool that is widely used for TD as well as autistic populations at various ages across several languages (Chen et al., 2023; Hutchins et al., 2012). Language proficiency in two languages is measured through the detailed parental questionnaire Quantifying Bilingual Experience (De Cat et al., 2023). Further, the study is the first to investigate the correlation between proficiency in a bilingual's stronger and weaker languages and questionnaire-based ToM at two time points, assessing a subset of participants approximately 1 year later. The secondary goal of this work is thus to determine whether the links observed in the initial testing hold true in the follow-up assessment.

Theory of mind and language in typically developing and autistic children

Theory of Mind refers to a set of interrelated abilities central to understanding and navigating the social world. First, it entails recognizing that others possess mental states such as thoughts, beliefs, desires, intentions, and emotions that differ from one's own (Baron-Cohen, 1995). This diversity highlights ToM as a multifaceted construct rather than a unitary ability (Tager-Flusberg, 2000). Individuals with strong ToM skills tend to excel in social communication, enabling them to build and maintain interpersonal relationships and function effectively in social contexts (Astington et al., 2002).

Typically developing children begin to distinguish between mental states and physical reality by age two, recognizing that people feel happy when desires are fulfilled and sad when they are not (Astington & Jenkins, 1999). By age 3, TD children begin to understand diverse beliefs and demonstrate competence in knowledge access tasks. Between ages 4 and 5, basic ToM abilities consolidate, enabling success in first-order false belief tasks. They also begin to understand the emotional consequences of beliefs, typically mastering related tasks between ages 5 and 6 (Wellman & Liu, 2004). At around 6 or 7, TD children start comprehending complex mental states such as understanding that an individual may attribute a false belief to another person, which itself may be inaccurate (Astington et al., 2002). These skills continue to mature into adolescence, supporting empathy, perspective-taking, and social reasoning, as shown in a meta-analysis by Qiu et al. (2024). Autistic children often show delays or impairments in ToM development

(Baron-Cohen, 1995). Many struggle with tasks that require understanding others' beliefs, intentions, or emotions, even when matched with TD peers on age or IQ (Baumeister et al., 2024).

Studies have consistently shown that EFs and first language development play a crucial role in the emergence and refinement of ToM in children (Pellicano, 2007; Schaeffer et al., 2021). Some research suggests that the development of ToM is contingent on EF abilities (Pellicano, 2010), whereas others contend that acquiring ToM skills comes first and may support the development of EFs (Perner & Lang, 1999).

Linguistic abilities such as lexical, grammatical, and syntactic skills were found to be important for ToM reasoning in monolingual TD and autistic children (De Villiers & Pyers, 2002; Durrleman et al., 2016; Farrar & Maag, 2002; Tager-Flusberg & Joseph, 2005). As for bilingual children, while some research also postulates a link between the first acquired language (L1) and ToM performance in TD (Białecka et al., 2024; Gordon, 2016) and autistic children (Peristeri et al., 2024), there is also evidence that metalinguistic awareness of bilinguals plays a primary role in ToM performance beyond their L1 language knowledge (Diaz & Farrar, 2018).

Theory of mind: behavioral tasks vs. informant questionnaires

While a great bulk of studies has measured ToM through various behavioral tasks, parent-based questionnaires were used less frequently, as shown in five current systematic reviews of ToM measures by Baumeister et al. (2024). One key challenge of measuring ToM through behavioral tasks is the disparity between the conceptualizations of ToM, encompassing constructs such as desires, beliefs, and emotions, as well as the limited scope of the tasks used to measure it, which may undermine the validity of score interpretations (Beaudoin et al., 2020). Further, numerous assessments are noted to impose substantial cognitive and linguistic demands on children, posing challenges for those on the autistic spectrum displaying cognitive or linguistic delays (Fu et al., 2023). Also, many behavioral ToM measures lack strong psychometric foundations, with only a limited number demonstrating acceptable levels of internal consistency (i.e., the degree to which different components within ToM tasks consistently assess the same construct), inter-rater reliability (i.e., the extent to which different observers produce similar scores for the same behavior), or criterion validity (i.e., how well a measure correlates with an established standard) (Ziatabar Ahmadi et al., 2015). Finally, behavioral ToM tasks performed in laboratory settings fall short in capturing how ToM skills are applied in everyday social contexts (Baumeister et al., 2024). Consequently, parental questionnaires have emerged as complementary tools for gauging a wider spectrum of ToM as it manifests in everyday life.

Parental questionnaires have been used together with behavioral tasks by several studies to measure ToM abilities in children. For instance, Mazzone and Nader-Grosbois (2016) examined how parents' reactions to their children's emotions relate to ToM abilities. Using questionnaires completed by 175 parents, they found that supportive maternal reactions to positive emotions were positively associated with children's ToM emotions and ToM thinking, while non-supportive reactions to negative emotions were linked to lower ToM performance.

Further, Houssa et al. (2013) investigated how short-term training in social cognition affects preschoolers' development of ToM and Social Information Processing (SIP) skills. Pre- and post-tests assessed emotion recognition, ToM, and social problem-solving. Parents completed the ToMI and a social adjustment scale while teachers filled out a personality questionnaire and the Child Behavior Checklist. At pretest, higher ToMI scores were positively correlated with better performance on direct ToM and social cognition tasks. This suggests that parents' perceptions of their children's ToM abilities aligned with objective measures. Also, children with stronger social cognition skills (as measured by ToM and SIP tasks) were rated by teachers as having better social adjustment and fewer externalizing behaviors (e.g., aggression, impulsiveness).

Although there are some limitations to using questionnaires in research, such as potential bias related to accuracy or trustworthiness (Zhang & Aryadoust, 2022), parental questionnaires have nonetheless proven to be a useful tool for atypical populations such as autistic children. For instance, parental assessments enable the evaluation of a wider range of constructs, providing a more comprehensive overview of children's ToM abilities in everyday settings.

Theory of mind and bilingualism in typically developing and autistic children

Studies on the link between bilingualism and ToM in TD children generally report that bilingual children outperform monolingual peers (Schroeder, 2018). That may be due to various reasons such as greater metalinguistic awareness (Doherty, 2000) or better executive control abilities compared to monolingual speakers (Bialystok, 2011). A meta-analysis by Schroeder (2018) investigated whether bilingual children outperformed monolingual peers on ToM tasks. The initial data analysis of 16 studies involving 1,283 children revealed a small but statistically significant bilingual advantage in ToM. Further, a secondary data analysis was conducted using a subset of eight studies in which differences in language abilities between bilingual and monolingual children were controlled. This follow-up analysis aimed to isolate the effect of bilingualism itself on ToM performance, independent of disparities in language proficiency. Notably, the results of the secondary analysis indicated a stronger correlation between bilingualism and ToM compared to the initial analysis.

A scoping review by Baumeister et al. (2025) summarized 37 studies that investigated the relation between bilingualism and ToM. Most of the reviewed studies investigated TD children, while only four studies looked at autistic children. Further, most studies compared monolingual speakers vs. bilingual speakers on a categorical basis, namely with a bilingual criterion predetermined before the study.

The results for TD children were mixed: about half of the studies found that bilingual children outperformed their monolingual peers; around 20% of the studies showed that monolingual and bilingual children did not differ from each other; approximately 25% of the studies reported mixed results depending on the exact ToM task (e.g., first- or second-order ToM); and only one study found a negative relation between bilingualism and ToM. Out of the four studies investigating the link between bilingualism and ToM in autistic children, three found a bilingual advantage

(Andreou et al., 2020; Peristeri et al., 2021a, 2021b) while one study reports no differences between bilingual and monolingual autistic children (Meir & Novogrodsky, 2019). However, there are also other studies on bilingualism and ToM in autistic children which were not mentioned in this review, revealing that bilingual autistic children outperform their monolingual peers (Baldimtsi et al., 2020; Peristeri et al., 2024). Additionally, most of the studies that found a positive correlation between bilingualism and ToM in TD children investigated children in the age range between 3 and 6 years old performing ToM behavioral tasks (Baumeister et al., 2025).

Studies that showed a positive relation between bilingualism and ToM compared monolingual to bilingual speakers using different bilingual predictors. For instance, some studies included the use of a weaker language and the amount or richness of exposure to it, while other studies investigated proficiency in the weaker language or the age of first exposure, as shown in a scoping review by Feng et al. (2024). Importantly, studies that found positive outcomes between bilingualism and ToM often examined participants with *significant* language exposure (Baumeister et al., 2025; Ricciardelli, 1992). This is important since some research shows that children must reach a certain level of proficiency in both languages to show cognitive advantages over their monolingual peers, as predicted by the Threshold Hypothesis (Cummins, 1976).

Among various bilingual variables, proficiency in the second language (L2) has been found to be a robust predictor of better performance on EFs, such as cognitive control, conflict monitoring (Xie, 2018) and cognitive flexibility (Yang et al., 2021) as well as on ToM tasks (Białecka et al., 2024; Gordon, 2016). One of the recent studies by Białecka et al. (2024) investigated the association between proficiency in sequentially first acquired language (L1), different bilingual predictors and ToM that was assessed through various tasks (e.g., first-order false-belief tasks, second-order false-belief, understanding of deception, ambiguity, and surprise). Participants were TD Polish-English bilingual migrant children residing in the UK and Polish monolingual children aged 4 to 6. The bilingual predictors examined were notably *L2 proficiency in English*, *L2 length of exposure*, *cumulative language exposure to L1 Polish*, and *cumulative language exposure to L2*. The study found that L1 proficiency and age were significant predictors of ToM accuracy for both monolingual (L1 English) and bilingual children (L1 Polish). Further, L1 proficiency was important for monolingual children when they provided justification for ToM answers (i.e., children had to explain *why* they answered the way they did). Interestingly, among bilingual children, both their L1 and L2 proficiency emerged as significant predictors of the justification index even though the task did not require the children to use their L2 English. Other bilingual predictors did not play a role for the accuracy or the justification index of ToM. Finally, although bilingual and monolingual children behaved alike regarding accuracy in ToM tasks, bilingual children performed better than their monolingual peers when they provided justification for ToM answers.

Zooming further into the link between language proficiency and ToM, Białecka et al. (2024) summarized the results of 13 studies on TD children which investigated the correlation between language proficiency (L1, L2 or both) and ToM, reporting a bilingual advantage in most of these studies ($N = 8$). In most of all the other studies that did not find a bilingual advantage, language abilities of monolinguals were higher than those of bilinguals. Overall, a ToM advantage was evident when language abilities were fully controlled, or when bilingual participants had L2 proficiency comparable to or greater than that of monolinguals—particularly when the testing was conducted in L2.

The correlation between language proficiency and ToM across *two* languages of autistic children is unknown. The existing published studies that examined the link between bilingualism and ToM in autism considered the following bilingual predictors: *current language use* quantified as a number of speakers interacting with the child (Andreou et al., 2020), *language preference* in everyday life with family members and friends (Peristeri et al., 2021a), *proficiency in L1* (Peristeri et al., 2021b, 2024), *age of onset* of bilingualism, *length of exposure* (months), and *current exposure to the dominant language* (Meir & Novogrodsky, 2019). Finally, other studies did not examine specific variables, since bilinguals were defined as speakers that were simply exposed to another language since birth (Baldimtsi et al., 2020).

Longitudinal studies on bilingualism and ToM in TD and autistic children are scarce. There is evidence that monolingual 3–4 years old TD children showed a significant improvement in their ToM performance between the first and the second testing point, while bilingual children did not show such improvement over time (Goetz, 2003). Further, bilingual children outperformed their monolingual peers on ToM performance at the first testing point but not at the second one. It has to be noted, though, that there was only a 1-week interval between the two testing sessions in the study by Goetz (2003). Therefore, this study examined short-term effects of bilingualism on ToM, leaving the question of long-term effects open. Research also found that metalinguistic awareness, rather than language knowledge (indexed as L1 or L2 proficiency) was a predictor of ToM performance in bilingual children at the second testing time point (Diaz & Farrar, 2018). The children were first tested at around age of 4 and approximately 1 year later. The only longitudinal study examining the relation between bilingualism and ToM in autistic children known to the authors is the one by Peristeri et al. (2024), which examined the link between bilingualism and the first- and second-order ToM skills of monolingual and bilingual autistic children at three points, namely at ages 6, 9, and 12. The findings indicate that bilingualism was associated with second-order ToM abilities in autistic children beyond the age of 9, with this advantage largely attributable to their expressive vocabulary in the societal language and intelligence.

Taken together, the reviewed accounts suggest that bilinguals' language skills (e.g., metalinguistic awareness, expressive vocabulary, and the use of language as an explicit support for social reasoning) facilitate ToM, thereby motivating the focus on language proficiency as the key predictor in the present study.

The present study

The current study aims to address gaps in the literature by examining how the language of higher proficiency and the language of lower proficiency correlate with ToM development in TD and autistic children over time. The current study is the first longitudinal study on the link between bilingualism and daily-life ToM skills. Furthermore, the current study solely focuses on bilingual speakers, since it seeks to determine whether the *amount* of proficiency in two languages matters for ToM.

The study is confirmatory oriented (language proficiency and ToM was investigated for TD children) with exploratory extensions (i.e., no previous longitudinal studies on proficiency in two languages and questionnaire-based ToM in autistic children). It aims at answering the following research questions (RQs): **RQ1**: What is the link between the language of higher proficiency (LHP), the language of lower proficiency (LLP) and questionnaire-based ToM in bilingual autistic and TD children? **Hypothesis 1**: Higher proficiency in both languages will be associated with better ToM performance in both autistic and TD bilingual children, following previous research in autistic (Baumeister et al., 2025) and TD children (Baumeister et al., 2025; Białecka et al., 2024). **RQ2**: Do the associations between LHP, LLP, and questionnaire-based ToM hold at the second testing time point (T2) a year later? Since there are no previous longitudinal studies that examine proficiency in two languages and ToM in TD and autistic children over time, this question remains to be explored in the current study. Therefore, we refrain from formulating specific hypotheses for this research question.

Method

Participants

The overall sample for the current study at the first testing point (further T1) consists of 244 bilingual, White children: 137 TD children (mean age in years = 9.2, range: 6.1–11.9) and 107 autistic children (mean age in years = 9.1, range: 6.1–12.0). The T1 dataset stems from a larger dataset from a previous cross-sectional study (Baumeister et al., 2026). A subset of the participants was tested approximately 1 year later. The sample for the second testing point (further T2) consists of 49 children: 28 TD children (mean age in years at T1 = 9.1, range: 6.1–11.5; mean age in years at T2 = 10.2, range: 7.1–12.6) and 21 autistic children (mean age in years at T1 = 9.2, range: 6.3–11.6; mean age in years at T2 = 10.1, range: 7.5–12.6). The children in the subset tested at T2 were all the children who agreed to participate in the study the second time. Furthermore, children were required to be bilingual with measurable proficiency in two languages, to fall within the specified age range, and (for the autistic group) to have a prior clinical diagnosis reported by caregivers. The data collection (T1/T2) took place from January 2023 to December 2024.

Children were diagnosed with ASD by a certified clinician, supported by either the Autism Diagnostic Observation Schedule-2nd Edition (ADOS-2) (Lord et al., 2012) or another standardized ASD diagnosis tool such as the Autism Diagnostic Interview-Revised (Lord et al., 1994). The diagnoses were provided by the parents. According to caregivers, typically developing children did not have a history of developmental disorders or language or cognitive delays. Although every parent confirmed that their child had an official diagnosis, not every parent provided an official document due to some privacy issues. Parents further filled out a Social Communication Questionnaire (SCQ) that is a widely used screening tool for identifying symptoms of ASD (Rutter et al., 2003). A score of 15 or higher suggests that an individual may be on the autism spectrum and should receive a full diagnostic evaluation. We are aware that these cutoffs are guidelines and not diagnostic. However, the SCQ scores are nevertheless informative since they provide an indication of autism severity. Autistic children in our overall data set had an average SCQ score of 19.2, while TD children had an average score of 4.8. TD children showed low SCQ scores below screening thresholds, reducing the risk of undiagnosed autism.

Prior to testing, parents provided their written consent to allow their children to participate in the study. All participants received gift cards as a means of compensation for their time and investment in the study. Socioeconomic status (SES) was determined based on the highest educational attainment of a primary caregiver, rated on a 5-point scale ranging from 1 (no education beyond primary school) to 5 (university-level education) (Hauser & Warren, 1997). SES values in our overall sample were generally high (between 4 and 5), suggesting that the sample predominantly came from households with substantial formal education.

Participants were recruited in different countries (i.e., Switzerland, France, Canada, Scotland, Spain, the USA, and Germany) through various channels such as autism associations, schools, social networks (Facebook), speech and language therapists, and the official US recruitment platform "BuildClinical." Such multi-channel and international recruitment strategy allowed us to examine children with diverse language backgrounds and language proficiencies. Although such cross-country recruitment can introduce further variability due to differences in diagnostic practices and prevalence rates, we applied consistent inclusion criteria across sites and verified diagnoses using standardized measures where available.

In this study, we present the data of the participants who were to some extent proficient in the two languages. The study deliberately does not refer to the two languages of children as L1 or L2 due to a heterogeneous data sample. Some children in our data set acquired their L1 which was also the societal language (mainly French, German, Spanish, or English), and they first started learning their L2 language in the formal settings in school. However, the first language of some other children was their home language and not the language of the larger society. These children are typically classified as heritage speakers, whose language dominance and proficiency often shifts toward the societal language with the start of schooling. Heritage speakers vary widely in their proficiency, i.e., some are fluent in their home language, while others may only understand it passively. Their home language input is usually informal and family based, whereas the societal language dominates formal and educational contexts (Wiese et al., 2025). Further, since our participants were aged 6 to 12 (mean age 9), it is likely that exposure to and dominance in the home language was declining, while the societal language was becoming increasingly dominant. Given these facts, labeling a home language of children as their L1 would not accurately reflect their current language proficiency in that language and would not be comparable to other children who acquired their L1 that was also a societal language and who did not experience diminishing input in that language. Therefore, to better reflect our diverse sample, we adopted an approach that describes two languages of children as more or less proficient rather than L1 and L2.

According to the Quantifying Bilingual Experience questionnaires that were filled out by children's parents, there was a wide range of LHP (13 languages) in the overall data set. The most frequent languages were French, German, Spanish, and English comprising around 80% of the LHP in the TD group and up to 93% in autistic group. The range of LLP in the full data set was even wider than that of LHP (34 languages), and English, Spanish, German, French, Catalan, and Italian were the most widespread LLP in our data set comprising around 70% of all LLP (see Online Material Table 1 for an overview of the diversity of LHP and LLP). Further, 26% of children ($N = 63$) acquired their LHP as their

Table 1 Participants' biographic characteristics: all participants tested at T1.

	All participants tested at T1	
	ASD	**TD**
	(*N*=107)	**(*N*=137)**
Age (years)	9.09 (1.68) [6.05–11.97]	9.17 (1.69) [6.07–11.93]
Sex (at birth)		
Female *N* (proportion)	24 (22.4%)	74 (54.0%)
Male *N* (proportion)	83 (77.6%)	62 (45.3%)
Prefer-not-to-specify	0 (0%)	1 (0.7%)
SES	4.18 (0.98) [1.00–5.00]	4.68 (0.71) [2.00–5.00]
IQ	93.52 (13.89) [66.00–128.00]	99.98 (12.69) [71.00–129.00]
SCQ score	19.20 (7.07) [1.00–34.00]	4.80 (3.52) [0.00–23.00]
LHP	0.73 (0.22) [0.17–1.00]	0.89 (0.14) [0.33–1.00]
LLP	0.49 (0.25) [0.08–1.00]	0.63 (0.21) [0.08–1.00]

Note. LHP = Language of higher proficiency; LLP = Language of lower proficiency; SCQ = Social Communication Questionnaire; SES = socioeconomic status.

sequentially first language, 32% of children (*N*=78) acquired their LHP as their sequentially second language while 42% of children (*N*=103) acquired their LHP as their sequentially first language but they also started simultaneously acquiring another language since birth at home.

Additionally, we analyzed non-verbal IQ of those participants who completed the test (i.e., 113 TD children and 82 autistic children) (Ravens-2; Raven et al., 2018). Both groups of children in the overall data set had average IQ scores (mean IQ for autistic children: 93.5; TD children: 100). An overview of participant characteristics in the overall data set is given in Table 1.

Since the subset of children tested at T2 (*N*=49) was much smaller than the one that included all children tested at T1 (*N*=244) it is important to present the participants' characteristics to see whether a small subset considerably differs from the children that were not tested at T2 but were tested at T1 only (*N*=195). Overall, the smaller sample tested at T2 was quite similar to the participants who were only tested at T1. Specifically, the subgroup of participants at T2 and those not tested at T2 come from households with generally high SES (between 4 and 5), both samples have average IQ values in the range of 90.5–101, SCQ scores for TD children are below 5 and in the range of 17.9–19.5 for autistic children. The most frequently spoken LHP are also French, German, Spanish, and English for both data sets (see a detailed description in Table 2).

Since the subset of participants at T2 is considerably smaller than that tested at T1, performing the analysis only on this small subset would yield a result that may not generalize well to the complete data. Therefore, the current study will present the data of the overall set of participants available at T1 as well as those tested at T2. An overview of participant characteristics at T2 is given in Table 2.

Materials

Quantifying Bilingual Experience questionnaire

To quantify the bilingual experience of the participants, their parents filled out the online Quantifying Bilingual Experience (Q-BEx) questionnaire (De Cat et al., 2023). Parental questionnaires are reported to reliably predict the input and experience of bilingual speakers, as seen in a review of 39 studies by Paradis (2017). The Q-BEx questionnaire is comprised of different modules that aim to elicit background information as well as information on language exposure and use, language proficiency, richness of linguistic experience, attitudes and satisfaction with child's language, and language mixing. Caregivers required approximately 1 hour to finish the questionnaire. Most of the caregivers filled the Q-BEx out only once, at T1, but a few of them filled it out at T2 as well, when caregivers indicated a change in the linguistic environment of the child (e.g., the family moved abroad, the child was exposed to a new language at school, etc.) that resulted in a change of proficiency in any of their languages. Out of 49 children tested in this study at T2, 28 caregivers filled the Q-BEx out at T2 as well.

In this study, we concentrate on the data from background information and proficiency modules. Language proficiency was calculated as a composite score that was comprised of four questions that assessed participants' proficiency in *speaking, understanding, writing,* and *reading* in each language to which the child was exposed (e.g., "how well can your child [speak] in [language]?"). Each domain was rated on a 4-point Likert scale ("hardly at all" to "very well"), further converted to scores from minimum 0 to maximum 3. The *relative average proficiency* score was calculated by averaging the scores from all four domains and dividing that score by 3 since that was the maximum rated proficiency. Each participant received one relative proficiency score in the range from 0 till 1, where the closer the number is to 1, the higher the rated proficiency for that language. Some children were quite balanced regarding their current proficiency in their LHP and LLP. Specifically, 15% of children in our data had either the same level of proficiency in both languages or they were similarly proficient in both languages, with less than a 10% difference between them. In these cases, the LHP was determined based on the country or specific region of testing (e.g., French for children tested in France; German for those tested in the German-speaking regions of Switzerland). Overall, TD children had higher proficiency than autistic children in both their LHP and LLP. As for the 28 children who had a proficiency change in their LHP or LLP, both TD and autistic children had an increase of their LLP by T2. As for the LHP, while autistic children showed an increase of proficiency, TD children showed a very slight negative change in their proficiency. Proficiency change was considered when the values for LHP or LLP at T2 changed compared to the values at T1.

The Q-BEx enabled us to estimate LHP and LLP in children across a wide range of languages, which would not be feasible with a test targeting a single language (e.g., testing lexicon). An overview of relative language proficiency scores in LHP and LLP across two time points is given in Table 2.

Theory of mind inventory 2 (ToMI-2)

The Theory of Mind Inventory-2 (ToMI-2) (Hutchins et al., 2012) is a parent-report questionnaire that assesses children's abilities of ToM in everyday life situations. ToMI-2 is a validated assessment

Table 2 Participants' biographic characteristics in the small data set.

	Participants tested at T1 only		Participants tested at T1 and T2	
	ASD	**TD**	**ASD**	**TD**
	(*N* = 86) M (SD) [range]	**(*N* = 109) M (SD) [range]**	**(*N* = 21) M (SD) [range]**	**(*N* = 28) M (SD) [range]**
Age at T1	9.07 (1.68) [6.05–11.97]	9.19 (1.70) [6.09–11.93]	9.19 (1.68) [6.33–11.62]	9.09 (1.66) [6.07–11.54]
Age at T2			10.11 (1.70) [7.46–12.65]	10.15 (1.66) [7.12–12.59]
Sex (at birth)				
Female *N* (proportion)	19 (22.1%)	59 (54.1%)	5 (23.8%)	15 (53.6%)
Male *N* (proportion)	67 (77.9%)	50 (45.9%)	16 (76.2%)	12 (42.8%)
Prefer-not-to-specify	0 (0%)	0 (0%)	0 (0%)	1 (3.6%)
SES at T1	4.08 (1.00) [1.00–5.00]	4.68 (0.67) [2.00–5.00]	4.53 (0.84) [2.00–5.00]	4.68 (0.85) [2.00–5.00]
SES at T2			4.57 (0.81) [2.00–5.00]	4.71 (0.81) [2.00–5.00]
IQ at T1	94.44 (13.87) [66.00–128.00]	99.68 (12.47) [71.00–129.00]	90.47 (13.89) [69.00–126.00]	101.04 (13.64) [71.00–122.00]
IQ at T2			85.46 (9.16) [69.00–99.00]	105.42 (14.63) [76.00–150.00]
SCQ score at T1	19.51 (7.13) [3.00–34.00]	4.91 (3.55) [0.00–23.00]	17.90 (6.83) [1.00–29.00]	4.36 (3.42) [0.00–13.00]
LHP **at T1**	0.73 (0.23) [0.17–1.00]	0.89 (0.14) [0.33–1.00]	0.69 (0.22) [0.33–1.00]	0.91 (0.14) [0.50–1.00]
LHP at T2			0.72 (0.21) [0.42–1.00]	0.90 (0.16) [0.42–1.00]
LLP at T1	0.50 (0.26) [0.08–1.00]	0.64 (0.22) [0.08–1.00]	0.41 (0.19) [0.08–0.75]	0.60 (0.18) [0.25–0.92]
LLP at T2			0.47 (0.22) [0.08–0.83]	0.74 (0.18) [0.25–1.00]

Note. LHP = Language of higher proficiency; LLP = Language of lower proficiency; SCQ = Social Communication Questionnaire; SES = socioeconomic status.

tool that is used for TD as well as autistic populations at various ages. The main strengths of ToMI-2 lie in its ability to tap into real-world every day naturalistic scenarios and in its ability to reduce high load on language and EFs typical for task-based ToM assessments. These strengths of ToMI-2 make it particularly suitable for autistic populations who may not perform reliably on structural behavioral tasks (Chen et al., 2023 for a successful implementation of ToMI-2 in autism), as well as for younger children and participants with heterogeneous language profiles.

The ToMI-2 is comprised of 60 statements addressing the child's ToM abilities in everyday life (e.g., "*My child is able to put himself or herself in other people's shoes and understand how they feel*" or "*My child knows how to make up stories to get what he or she wants*"). Caregivers were asked to assess to which extent each statement applies to their child on a continuum from "definitely not" to "definitely." Each of the 60 ToMI-2 items was rated by caregivers on a 0–20 scale. For each subscale, the 20 item ratings were summed and divided by 20 to produce a subscale score ranging from 0 to 20. The overall ToMI-2 score was calculated by summing all 60 items and dividing by 60, yielding an overall score also ranging from 0 to 20. The higher the score, the more developed the child's ToM skills.

The online platform Gorilla Experiment Builder was used to administer ToMI-2 (Anwyl-Irvine et al., 2020). The ToMI-2 was administered in either English, French, German, Italian or Spanish. Further, we corrected some items in the translations, as they contained mistakes. All translations were published and can be found here: https://osf.io/r4s3c/overview?view_only=067873d8f0d645f89dc683e33d24a03a. Three empirically derived ToMI sub-scales (i.e., early, basic, and advanced) reflect a developmental progression in ToM development. The early subscale (e.g., emotion and desire recognition) tests abilities that are achieved in infancy and toddlerhood and it corresponds to preschool tasks described by Wellman and Liu (2004). The basic subscale (e.g., false belief and perspective-taking) tests abilities that develop around and after 4 years of age, and it aligns with classic false belief tasks such as the Sally-Anne test (Baron-Cohen et al., 1985). The advanced subscale (e.g., sarcasm, irony, second-order beliefs) corresponds to abilities that emerge later in middle childhood and adolescence (from 7), and it maps onto later-developing abilities studied by Perner and Wimmer (1985) (e.g., Ice Cream Truck Story task). Unlike studies that use single tasks at specific ages, ToMI-2 integrates these into a composite measure spanning early childhood through adolescence. ToMI-2 provides both composite (across all three subscales) and subscale scores (a score per subscale). This study examines composite ToMI-2 scores from three abovementioned subscales (average score of all ToMI-2 questions; minimum 0 to maximum 20).

Data analysis

The main goal of this study is to investigate relations between the two languages (LHP and LLP) and ToM in bilingual TD and autistic children, and to see whether the results at T1 hold longitudinally at T2 as well. To achieve this, a linear regression model was fit to all available T1 data, and performance (residuals of the predictions and errors) of this very model was checked and compared between T1 and T2.

We considered using other statistical methods, but these were ruled out either because of not leveraging the complete data set, not fulfilling the statistical preconditions needed for the application, difficulties related to the low number of samples, and/or difficulties regarding the model interpretation. The following options were considered for data analysis:

1. A single mixed effect model for all data: to perform statistical analysis on repeated measures, a linear mixed-effect model with a random effect for the participant could be used, modelling T1 and T2 both. In the dataset, only a very small subset (20.1%) of the participants was available for a longitudinal analysis (49 participants tested at T1 and T2 vs. 244 participants tested at T1 only): therefore, most participants had only one single data point. There are no clear recommendations for cases like these with respect to choices regarding the random effects structure, i.e., whether adding a random effect for participants with a single data point is prone to errors or not (it is possible if there is at least 30% of data with multiple data points (Bell et al., 2010), also pooling of these participants was considered, but that is not a real option as it will distort the results.
2. Analysis of only a small subset that had T1, T2 data: this would have meant losing a substantial proportion of the data from the first time point, resulting in model(s) not generalizing well to the larger dataset (overfit).
3. Leverage the T1 of data for determining priors for the T1-T2 in Bayesian modelling. Setting strong priors based on T1 would influence or limit the outcome of the T1-T2 modelling that adds only a little data.

After ruling out the alternative methods, we pursued the data analysis for the complete available dataset in a way that avoids overfitting, leverages the full data set, and allows us to manage the low number of samples at T2.

Step 1: linear modeling of the data at T1

For statistical analysis, the linear regression model (see Online Material, Table S2 for the exact model formula) was created in R for all the participants at T1 (R Core Team, 2020). The linear model models the average ToMI-2 scores as a function of *group* (TD or ASD) and *age* and includes two-way interactions of *group* with two predictors of the main focus in this study: *LHP* and *LLP*. The linear fixed effects were scaled by the R scale command. The by-participant individual ToMI-2 score is an average of all the questions in the ToMI-2 questionnaire. The individual ToMI-2 questionnaire ratings have a special distribution: since the range is limited to 0–20, the scores of 0, 10 and 20 were frequently selected, especially saturated for many TD children at the maximum, therefore modelling the individual questions' scores would be challenging. It was possible to model the average ToMI-2 score despite some limitations. An exploratory data analysis guided the selection of an age range in which sufficient data were available, the ToMI score had not yet reached maximum saturation, and the contributing parameters exhibited a linear trend. This range was from 72 to 144 months. Model diagnostics with the DHARMa package via QQ plot of residuals and dispersion test revealed no significant issues with this model (Hartig, 2024). Correlation between LHP, LLP and multicollinearity were checked with the car (Fox & Weisberg, 2019) and mctest (Ullah et al., 2019) packages. The correlation between LHP and LLP was moderate for TD children (0.38), and it was strong for autistic children (0.65). Multicollinearity was not an issue in the model (values from 1.21 till 1.93). The R script and data file are available under: https://osf.io/psjue/

Step 2: prediction of the data at T2

To evaluate if the linear model can predict T2 performance well, the following error rates were calculated and compared: the root mean square error (RMSE) of the model at T1 data of those participants who had only a single time point; RMSE of the model at T1 of those participants who participated in T2; RMSE of the model at T2 (see Online Material, Model performance at T2 vs T1 for the exact formula). Furthermore, the residuals of the model at T1 (all data) were compared to the residuals at T2 via a Wilcoxon rank sum test.

Results

RQ1: correlation between the language of higher proficiency, the language of lower proficiency, and ToMI-2 at T1

According to descriptive statistics, TD children in the overall data set showed quite high ToMI scores at T1 (mean score: 17.3, $SD = 2.26$). Further, not a single TD child was at ceiling while around 30% scored near ceiling ($N = 42$, scores from 19.05 to 19.96). Autistic children showed ToMI scores that were lower than those of TD children (mean score: 11.3, $SD = 3.8$); not a single child one was at ceiling while one child scored near ceiling (i.e., 19.06).

The statistical model revealed that autistic children had significantly lower average ToMI-2 scores than TD children (*estimate* $= -2.64$, $SE = 0.19$, $t = -13.72$, $p < .001$). Further, there was a main effect of age showing that age positively predicted ToMI-2 in both groups (*estimate* $= 0.67$, $SE = 0.19$, $t = 3.51$, $p < .001$; see Figure 1 for the display of these main effects). A Type III ANOVA showed a significant group × LHP interaction ($F = 6.08$, $p = .003$), and a significant group × LLP interaction ($F = 3.14$, $p = .045$). Follow-up simple-slope analyses (emtrends) indicated that LHP positively predicted higher ToMI-2 scores in the autistic group (*slope* $= 1.049$, $SE = 0.3$, 95% CI [0.45, 1.65]) but not in the TD group (*slope* $= -0.15$, $SE = 0.38$, 95% CI [−0.9, 0.59]; see Figure 2). Further, LLP positively predicted higher ToMI-2 scores in the autistic group (*slope* $= 0.81$, $SE = 0.33$, 95% *CI* [0.16, 1.45]) but not in the TD group (*slope* $= 0.13$, $SE = 0.28$, 95% *CI* [−0.42, 0.68]; see Figure 3).

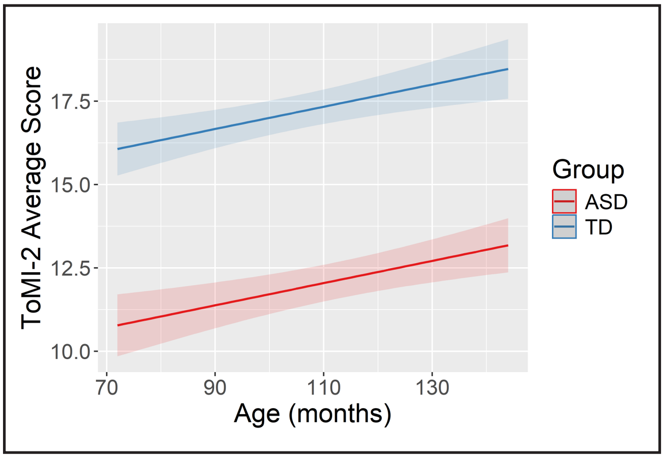

Figure 1 Associations between speaker group, age, and ToMI-2 scores in typically developing (TD) and autistic children (ASD).

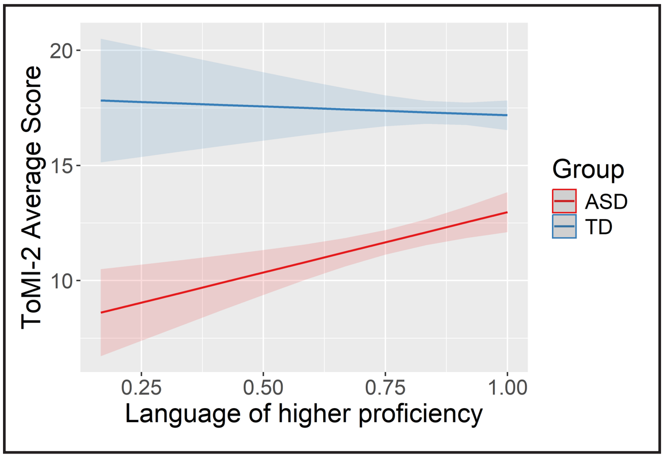

Figure 2 Associations between language of higher proficiency and ToMI-2 scores in typically developing (TD) and autistic children (ASD).

Overall, the results of T1 revealed that TD children clearly outperformed autistic children by showing significantly higher average ToMI-2 scores. Additionally, ToMI-2 scores improved with age for both groups. Further, LHP and LLP predicted higher ToMI-2 scores for autistic but not for TD children. In the next step, the linear model presented above was tested for T2.

RQ2: associations between the language of higher proficiency, the language of lower proficiency, and ToMI-2 at T2

According to descriptive statistics, TD children who were tested at T2 showed high ToMI scores at T1 (mean score: 17.6, $SD = 2.22$) slightly improving at T2 (mean score: 17.7, $SD = 2.33$). Further, one child scored at ceiling at T2 while nine children scored near ceiling (i.e., 19.1–19.92). Autistic children showed ToMI scores that were lower than those of TD children at T1 (mean score: 12.1, $SD = 4.11$) and T2 (mean score: 13.2, $SD = 3.95$), but the former also demonstrated a greater improvement in ToMI scores than the latter. Also, not a single autistic child was at ceiling at T2 while two children scored near ceiling (i.e., 19 and 19.04).

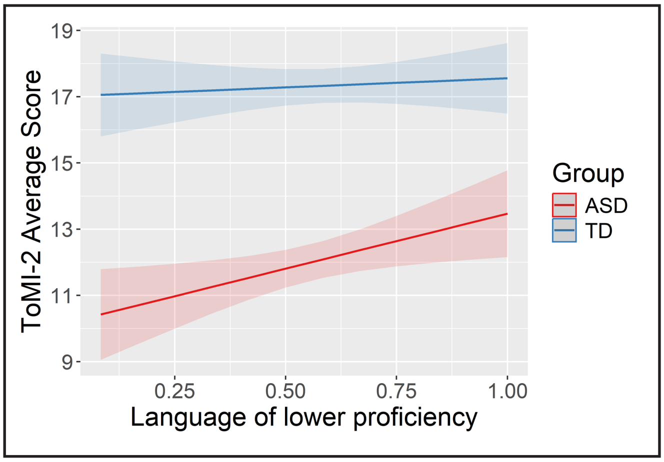

Figure 3 Associations between language of lower proficiency and ToMI-2 scores in typically developing (TD) and autistic children (ASD).

The residuals of the model at T1 (all data), compared to the residuals at T2 via a Wilcoxon rank sum test did not show a statistically significant difference ($W = 5302$, $p = .212$). The RMSE and R^2 values were very similar (i.e., $\text{RMSE} = 2.63$ and $R^2 = 0.624$ for participants at T1 who only have T1 data; $\text{RMSE} = 2.52$ and $R^2 = 0.628$ for participants at T1 who have T1 and T2 data; $\text{RMSE} = 2.59$ and $R^2 = 0.538$ for participants at T2 who have T1 and T2 data). Based on the RMSE and the comparison of the residuals, the model from T1 performed at T2 similarly to T1, indicating that the results of the linear regression model at T1 hold for T2 as well.

Discussion

The current study examined whether two languages of varying proficiency (i.e., LHP and LLP) correlate with daily-life ToM in bilingual autistic and TD children. Children were between 6 and 12 years of age at T1, and between 7 and 12.7 at T2. To our knowledge, it is the first longitudinal study that investigated ToM in daily life.

Our first research question focused on whether there was a link between LHP, LLP, and daily life ToM in TD children and autistic children at the first testing period. Our results only partially confirm Hypothesis 1, which predicted that higher proficiency in both languages would predict ToMI scores in both TD and autistic children. The results of the study showed that proficiency in LHP and LLP did not correlate with ToM in TD children. However, proficiency in LHP and LLP predicted ToMI-2 values of autistic children, demonstrating that higher proficiency of both languages was positively associated with higher ToMI-2 values.

The results of the current study contrast with the studies that found that LHP is crucial for ToM development, both for monolingual (De Villiers & Pyers, 2002; Farrar & Maag, 2002) and bilingual TD children (Białecka et al., 2024; Gordon, 2016). The possible discrepancy between the current study and the above-mentioned ones possibly lies in the age of children, their high ToMI scores, their high level of LHP, or the way language proficiency was measured. While the above-described studies examined younger TD children (i.e., preschoolers, ages around 3–5), the current study investigated TD children in the range from 6 until 12 years of age. TD children in our study showed high performance and low variability regarding their LHP (average of around and above 0.9 at T1, full and reduced data set, as well as at T2). Further, and very importantly, TD children in our data have already mastered their ToM to a high level reaching an average ToMI score above 17 at T1 and slightly improving at T2 to nearly 18 leaving very little room for further improvement and potential language effects.

These results are generally similar to the ones by Durrleman et al. (2016); Lind and Bowler (2009), and Tager-Flusberg (2000), who examined a relation between ToM and language-related skills, operationalized through syntactic tasks. Durrleman et al. (2016) found no association between non-verbal ToM and the understanding of complement clauses in TD children between 6.8 and 15.8 years old, while such an association was found for autistic children between 6.9 and 14.4 years of age. These findings also align with the study by Lind and Bowler (2009), who found an association between complement syntax competence and false belief task in autistic (mean age 10) but not in TD children of 9 years of age. These findings are also in line with Tager-Flusberg (2000), who did not find a relation between complement syntax and a first-order belief task in TD older children, but did in autistic children. Finally, the absence of correlations between language and ToM in TD children may not reflect a true lack of observable growth effects per se, but rather limitations of the Q-BEx questionnaire, which relies on ordinal parental ratings to assess general language development instead of capturing more fine-grained linguistic abilities.

Furthermore, this study also found that autistic children presented generally lower ToM skills than their TD peers, independently of their proficiency in two languages. This result is line with a number of studies on ToM in autism, notably showing that monolingual autistic children experience persistent ToM difficulties and rely longer on language than TD peers (Durrleman et al., 2016; Lind & Bowler, 2009; Tager-Flusberg & Joseph, 2005). For autistic children, language may serve as a helpful tool to overcome challenges in ToM by using language as an internal verbal dialogue, suggesting that an explicit linguistic rehearsal may help them to orientate through ToM tasks (Durrleman et al., 2016). TD children, on the other hand, might not need such facilitation through language since in TD children ToM difficulties generally diminish at around age of five. Thus, the absence of the association between the LHP and ToM in TD children in this study may be explained by their older age.

The older age of TD children might also explain the absence of the correlation between the LLP and ToMI-2 scores. Although some studies found a link between the LLP (that is referred to as L2 in some studies) and ToM, these studies examined younger children, typically preschoolers (Białecka et al., 2024; Gordon, 2016). Therefore, it is possible that TD children in our study reached an age at which neither their LHP nor their LLP played a pivotal role any longer for ToM development. Importantly, the current study reports on the novel findings, namely a correlation between the LLP and ToMI-2 scores in autistic children showing that not only proficiency in the LHP but also in the LLP relates to the improvement of ToM. Besides, not only the LHP and the LLP correlate with ToMI scores, but the LHP and the LLP showed a strong positive correlation (0.65) in autistic children suggesting that strong skills in one language are closely linked to strong skills in another language. This relation supports the developmental interdependence hypothesis (Cummins, 1979), which argues that proficiency in the LHP can positively influence proficiency in the LLP.

Furthermore, we conducted an exploratory analysis to examine whether the degree of dominance (i.e., the proficiency difference between the two languages) provides more explanatory power than absolute proficiency in each language. The results showed that it does not. Models based solely on dominance fit the data substantially worse than models that include absolute proficiency in each language. Moreover, the degree of dominance does not correlate with the ToMI scores in any of the speaker groups.

Further, we analyzed how strongly non-verbal IQ correlates with LHP and LLP in the autistic group. We found that IQ moderately correlates with both languages (LLP: 0.36, LHP: 0.45). This result implies that IQ explains some of the variation in language proficiency, but not all. Other factors also play a role. In addition, IQ has very little effect on the ToMI scores and does not significantly improve the model fit, as shown in our exploratory analysis.

While many studies highlight the importance of linguistic ability for ToM development, evidence also suggests the reverse effect: ToM can shape language growth. Longitudinal work has shown bidirectional links, with earlier ToM predicting later language and vice versa (see a meta-analysis by Milligan et al., 2007). To address the direction of the effect we attempted to model the rate of change between T1 and T2 for LHP, LLP, and ToMI-scores based on the predictors at T1 to see whether T2 can be predicted and some conclusions regarding causation can be drawn. Unfortunately, the linear regression models did not fit the data well, failing model fit tests most probably due to a limited data set at T2 (i.e., we had 21 autistic and 28 TD children tested at T2 and only half of them changed their language proficiency in the LHP or LLP). Therefore, it would be beneficial to have a greater data set to test the possible bidirectional links between ToM and language.

The results of this study regarding the relation between the LHP and ToM are in line with a longitudinal study by Peristeri et al. (2024) who reported a strong association between proficiency (as indexed by expressive vocabulary) in the societal language, Greek, and second-order ToM in bilingual autistic children at the age 9 and 12. However, it was unclear in that work whether there was any link between another language of bilingual children and ToM, since Peristeri et al. (2024) did not include language measures in a home language the children were exposed to since early childhood (i.e., Albanian). Therefore, the current study goes beyond Peristeri et al. (2024) by including the measures of proficiency in two of the children's languages.

Our second research question focused on whether the associations between LHP, LLP, and ToM hold across time. The current study found that proficiency in both languages correlates with ToMI-2 scores in autistic children not only at the first testing point, but also 1 year later as well. Our findings reveal that autistic children keep capitalizing on both languages *over time*. These results are relevant in light of the fact that parents of bilingual autistic children are sometimes advised against bilingual upbringing, due to fear that it could allegedly exacerbate challenges related to autism (Beauchamp & MacLeod, 2017; Hampton et al., 2017). The positive correlation between bilingualism and ToM in autistic children over time suggests that such advice is unfounded. Also, this study expanded the empirical understanding of proficiency in two languages and ToM in autistic children, which is essential to the autism community in making evidence-based decisions regarding multilingualism (Beauchamp & MacLeod, 2017).

Finally, recent theoretical critiques have emphasized that many commonly used ToM measures are limited in the extent to which they can be taken to index underlying cognitive mechanisms in autism (Long et al., 2025). In light of this, the present findings are best understood as characterizing associations between language proficiency and ToM as manifested in everyday social functioning, rather than as evidence bearing on specific mechanistic accounts. While future research should broaden longitudinal samples and integrate multiple approaches to assessing ToM, the present results nevertheless suggest that bilingual autistic children draw on both their stronger and weaker languages as stable resources for everyday mental state understanding over time.

Data availability

The data and code necessary to reproduce the analyses presented here are publicly accessible at the following URL: https://osf.io/psjue/?view_only=bb2ac2f35bca463697fc4a739a683c5c

The materials necessary to attempt to replicate the findings presented are available from the first author. The analyses presented here were not preregistered.

Author contributions

Yulia Zuban (Conceptualization, Writing—review & editing [equal], Data curation, Formal analysis, Investigation, Visualization, Writing—original draft [lead]), Franziska Baumeister (Data curation, Investigation, Methodology, Writing—review & editing [equal]), Pauline Lina Wolfer (Data curation, Investigation, Methodology, Writing—review & editing [equal]), and Stephanie Durrleman (Conceptualization, Methodology, Writing—review & editing [equal], Funding acquisition, Project administration, Resources, Supervision [lead])

Funding

This study was supported by the Swiss National Science Foundation (SNSF) awarded to Stephanie Durrleman (grant PR00P1_193104/1).*

Conflicts of interest

The authors declare none.

Supplementary material

Supplementary material is available at *Child Development* online.

Acknowledgments

We are grateful to Anna Czypionka for her constructive feedback.

Compliance with ethical standards and informed consent

This study was approved by the Swiss Association of Research Ethics Committees *Swissethics* (Project ID-2022-00878), the Institutional Review Board of the University of Connecticut (US), the Research Ethics Board Office at McGill University in Montreal (Canada), the Psychology Research Ethics Committee of the University of Edinburgh (UK), the Institutional Review Board of Emerson College Boston (US), and the Research Ethics Committee of the Unio Catalana Hospitals Foundation (Spain). All parents provided informed written consent for their child's participation prior to their inclusion in the study. All work has been carried out in accordance with The World Medical Association's Declaration of Helsinki for experiments involving humans.

References

American Psychiatric Association. (2022). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR)*. Washington, DC, USA: American Psychiatric Publishing.

Andreou, M., Tsimpli, I. M., Durrleman, S., & Peristeri, E. (2020). Theory of mind, executive functions, and syntax in bilingual children with autism spectrum disorder. *Languages*, *5*, 67. https://doi.org/10.3390/languages5040067

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388–407. https://doi.org/10.3758/s13428-019-01237-x

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, *35*, 1311–1320. https://doi.org/10.1037/0012-1649.35.5.1311

Astington, J. W., Pelletier, J., & Homer, B. (2002). Theory of mind and epistemological development: The relation between children's second-order false-belief understanding and their ability to reason about evidence. *New Ideas in Psychology*, *20*, 131–144. https://doi.org/10.1016/s0732-118x(02)00005-3

Baldimtsi, E., Peristeri, E., Tsimpli, I. M., & Durrleman, S. (2020). *The impact of bilingualism on theory of mind and executive functions in TD and ASD*. Proceedings of the 44th Boston University Conference on Language Development, 40–52. https://www.lingref.com/bucld/44/BUCLD44-04.pdf

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT Press. https://doi.org/10.7551/mitpress/4635.001.0001

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*, 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baumeister, F., Bagioka, D. V., Rivoletti, L., & Durrleman, S. (2025). The impact of bilingualism on theory of mind in children with and without developmental disorders: A scoping review. *Developmental Review*, *75*, e101186, https://doi.org/10.1016/j.dr.2025.101186

Baumeister, F., Wolfer, P., & Durrleman, S. (2026). Which aspect of the bilingual experience affects Theory of Mind in autistic and neurotypical children? *Research in Autism*, *133*, 202880. https://doi.org/10.1016/j.reia.2026.202880

Baumeister, F., Wolfer, P., Sahbaz, S., Rudelli, N., Capallera, M., Daum, M. M., Samson, A. C., Corrigan, G., Naigles, L., & Durrleman, S. (2024). Measuring theory of mind: A preliminary analysis of a novel linguistically simple and tablet-based measure for children. *Frontiers in Developmental Psychology*, *2*, e1445406, https://doi.org/10.3389/fdpys.2024.1445406

Beauchamp, M. L. H., & MacLeod, A. A. N. (2017). Bilingualism in children with autism spectrum disorder: Making evidence based recommendations. *Canadian Psychology/Psychologie Canadienne*, *58*, 250–262. https://doi.org/10.1037/cap0000122

Beaudoin, C., Leblanc, É, Gagner, C., & Beauchamp, M. H. (2020). Systematic review and inventory of theory of mind measures for young children. *Frontiers in Psychology*, *10*, 2905. https://doi.org/10.3389/fpsyg.2019.02905

Bell, B. A., Morgan, G. B., Kromrey, J. D., & Ferron, J. M. (2010). *The impact of small cluster size on multilevel models: A Monte Carlo examination of two-level models with binary and continuous predictors*. Proceedings of the Joint Statistical Meetings, Section on Survey Research Methods. 4057–4067. Alexandria, VA: American Statistical Association. http://www.asasrms.org/Proceedings/y2010/Files/308112_60089.pdf

Białecka, M., Wodniecka, Z., Muszyńska, K., Szpak, M., & Haman, E. (2024). Both L1 and L2 proficiency impact ToM reasoning in children aged 4 to 6. Painting a more nuanced picture of the relation between bilingualism and ToM. *Bilingualism: Language and Cognition*, *27*, 751–751. https://doi.org/10.1017/s136672892300072x

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology*, *65*, 229–235. https://doi.org/10.1037/a0025406

Chen, K.-L., Jiang, D.-R., Yu, Y.-T., & Lee, Y.-C. (2023). Development and psychometric evidence of the Chinese version of the theory of mind inventory-2 (ToMI-2) in children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *103*, 102132, https://doi.org/10.1016/j.rasd.2023.102132

Cummins, J. (1976). The influence of bilingualism on cognitive growth: A synthesis of research findings and explanatory hypotheses. *Working Papers on Bilingualism*, *9*, 1–43.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*, 222–251. https://doi.org/10.3102/00346543049002222

De Bruin, A. (2019). Not all bilinguals are the same: A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences*, *9*, 33. https://doi.org/10.3390/bs9030033

De Cat, C., Kašćelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S.; & The Q-BEx Consortium. (2023). How to quantify bilingual experience? Findings from a delphi consensus survey. *Bilingualism: Language and Cognition*, *26*, 112–124. https://doi.org/10.1017/S1366728922000359

De Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development*, *17*, 1037–1060. https://doi.org/10.1016/s0885-2014(02)00073-4

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diaz, V., & Farrar, M. J. (2018). The missing explanation of the false-belief advantage in bilingual children: A longitudinal study. *Developmental Science*, *21*, e12594. https://doi.org/10.1111/desc.12594

Doherty, M. J. (2000). Children's understanding of homonymy: Metalinguistic awareness and false belief. *Journal of Child Language*, *27*, 367–392. https://doi.org/10.1017/s0305000900004153

Durrleman, S., Burnel, M., Thommen, E., Foudon, N., Sonié, S., Reboul, A., & Fourneret, P. (2016). The language cognition interface in ASD: Complement sentences and false belief reasoning. *Research in Autism Spectrum Disorders*, *21*, 109–120. https://doi.org/10.1016/j.rasd.2015.10.003

Farrar, M. J., & Maag, L. (2002). Early language development and the emergence of a theory of mind. *First Language*, *22*, 197–213. https://doi.org/10.1177/014272370202206504

Feng, J., Cho, S., & Luk, G. (2024). Assessing theory of mind in bilinguals: A scoping review on tasks and study designs. *Bilingualism: Language and Cognition*, *27*, 531–545. https://doi.org/10.1017/S1366728923000585

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage. https://www.john-fox.ca/Companion/

Fu, I. N., Chen, K. L., Liu, M. R., Jiang, D. R., Hsieh, C. L., & Lee, S. C. (2023). A systematic review of measures of theory of mind for children. *Developmental Review*, *67*, 101061. https://doi.org/10.1016/j.dr.2022.101061

Giovannoli, J., Martella, D., Federico, F., Pirchio, S., & Casagrande, M. (2020). The impact of bilingualism on executive functions in children and adolescents: A systematic review based on the PRISMA method. *Frontiers in Psychology*, *11*, 574789. https://doi.org/10.3389/fpsyg.2020.574789

Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition*, *6*, 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*, 407–424. https://doi.org/10.1017/s0305000915000276

Grosjean, F., Li, P., & Bialystok, E. (2013). *The psycholinguistics of bilingualism*. Wiley-Blackwell/John Wiley & Sons. https://download.e-bookshelf.de/download/0000/7526/81/L-G-0000752681-0002328279.pdf

Gunnerud, H. L., Ten Braak, D., Reikerås, E. K. L., Donolato, E., & Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin*, *146*, 1059–1083. https://doi.org/10.1037/bul0000301

Hampton, S., Rabagliati, H., Sorace, A., & Fletcher-Watson, S. (2017). Autism and bilingualism: A qualitative interview study of parents' perspectives and experiences. *Journal of Speech, Language, and Hearing Research*, *60*, 435–446. https://doi.org/10.1044/2016_JSLHR-L-15-0348

Hartig, F. (2024). *DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models*. Version 0.4.7. [*Computer software*]. https://github.com/florianhartig/DHARMa

Hauser, R. M., & Warren, J. R. (1997). Socioeconomic indexes for occupations: A review, update, and critique. *Sociological Methodology*, *27*, 177–298. https://doi.org/10.1111/1467-9531.271028

Houssa, M., Nader-Grosbois, N., & Jacobs, E. (2013). Experimental study of short-term training in social cognition in pre-schoolers. *Journal of Education and Training Studies*, *2*, 139–154. https://doi.org/10.11114/jets.v2i1.181

Hutchins, T. L., Prelock, P. A., & Bonazinga, L. (2012). Psychometric evaluation of the theory of mind inventory (ToMI): A study of typically developing children and children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *42*, 327–341. https://doi.org/10.1007/s10803-011-1244-7

Lind, S. E., & Bowler, D. M. (2009). Language and theory of mind in autism spectrum disorder: The relationship between complement syntax and false belief task performance. *Journal of Autism and Developmental Disorders*, *39*, 929–937. https://doi.org/10.1007/s10803-009-0702-y

Long, E. L., Catmur, C., & Bird, G. (2025). The theory of mind hypothesis of Autism: A critical evaluation of the status quo. *Psychological Review*. https://doi.org/10.1037/rev0000532. Advance online publication.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule (ADI-R, ADITM-R)* (2nd ed.). Western Psychological Services. https://www.research.chop.edu/car-autism-roadmap/autism-diagnostic-observation-schedule-2nd-edition-ados-2

Lord, C., Rutter, M., & LeCouteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disororders*, *24*, 659–685. https://doi.org/10.1007/BF02172145

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*, 1115–1146. https://doi.org/10.1177/0956797621993108

Mazzone, S., & Nader-Grosbois, N. (2016). How are parental reactions to children's emotions related to their theory of mind abilities? *Psychology*, *07*, 166–179. https://doi.org/10.4236/psych.2016.72019

Meir, N., & Novogrodsky, R. (2019). Prerequisites of third-person pronoun use in monolingual and bilingual children with autism and typical language development. *Frontiers in Psychology*, *10*, 2289. https://doi.org/10.3389/fpsyg.2019.02289

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*, 622–646. https://doi.org/10.1111/j.1467-8624.2007.01018.x

Paradis, J. (2017). Parent report data on input and experience reliably predict bilingual development and this is not trivial. *Bilingualism: Language and Cognition*, *20*, 27–28. https://doi.org/10.1017/s136672891600033x

Pellicano, E. (2007). Links between theory of mind and executive function in young children with autism: Clues to developmental primacy. *Developmental Psychology*, *43*, 974–990. https://doi.org/10.1037/0012-1649.43.4.974

Pellicano, E. (2010). Individual differences in executive function and central coherence predict developmental changes in theory of mind in autism. *Developmental Psychology*, *46*, 530–544. https://doi.org/10.1037/a0018287

Peristeri, E., Baldimtsi, E., Vogelzang, M., Tsimpli, I. M., & Durrleman, S. (2021a). The cognitive benefits of bilingualism in autism spectrum disorder: Is theory of mind boosted and by which underlying factors? *Autism Research*, *14*, 1695–1709. https://doi.org/10.1002/aur.2542

Peristeri, E., Tsimpli, I. M., & Durrleman, S. (2021b). *Positive effects of bilingualism on social cognition in autism spectrum disorder: A study of social exclusion and theory of mind*. Proceedings of the 45th Annual Boston University Conference on Language Development, 626–638. https://www.lingref.com/bucld/45/BUCLD45-47.pdf

Peristeri, E., Vogelzang, M., Tsimpli, I. M., & Durrleman, S. (2024). Bilingualism and second-order theory of mind development in autistic children over time: Longitudinal relations with language, executive functions, and intelligence. *Autism Research: Official Journal of the International Society for Autism Research*, *17*, 1818–1829. https://doi.org/10.1002/aur.3214

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, *3*, 337–344. https://doi.org/10.1016/S1364-6613(99)01362-5

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that…" attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, *39*, 437–471. https://doi.org/10.1016/0022-0965(85)90051-7

Qiu, X., Gao, M., Zhu, H., Li, W., & Jiang, R. (2024). Theory of mind, empathy, and prosocial behavior in children and adolescent: A meta-analysis. *Current Psychology*, *43*, 19690–19707. https://doi.org/10.1007/s12144-024-05762-7

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

Raven, J., Rust, J., Chan, F., & Zhou, X. (2018). *Raven's 2 progressive matrices, clinical edition (Raven's 2)*. Pearson.

R Core Team. (2020). *R: A language and environment for statistical computing*. [*Computer software*]. http://www.r-project.org

Ricciardelli, L. A. (1992). Bilingualism and cognitive development in relation to threshold theory. *Journal of Psycholinguistic Research*, *21*, 301–316. https://doi.org/10.1007/bf01067515

Rutter, M., Bailey, A., Berument, S. K., & Lord, C. (2003). *Social communication questionnaire: Manual*. Western Psychological Services.

Schaeffer, J., Durrleman, S., & Eigsti, I.-M. (2021). On links between language development and extralinguistic cognitive knowledge: What we can learn from autism. *Language Acquisition*, *28*, 1–5. https://doi.org/10.1080/10489223.2020.1769631

Schroeder, S. R. (2018). Do bilinguals have an advantage in theory of mind? A meta-analysis. *Frontiers in Communication*, *3*, 36. https://doi.org/10.3389/fcomm.2018.00036

Silleresi, S. (2023). *Developmental profiles in autism spectrum disorder: Theoretical and methodological implications* (Vol. *68*). John Benjamins Publishing Company. https://doi.org/10.1075/lald.68

Tager-Flusberg, H. (2000). Language and understanding minds: Connections in autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed., pp. 124–149). Oxford University Press. https://doi.org/10.1093/oso/9780198524465.003.0006

Tager-Flusberg, H., & Joseph, R. M. (2005). How language facilitates the acquisition of false-belief understanding in children with autism. In J. W. Astington, & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 298–318). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159912.003.0014

Ullah, M. I., Aslam, M., Altaf, S., & Ahmed, M. (2019). Some new diagnostics of multicollinearity in linear regression model. *Sains Malaysiana*, *48*, 2051–2060. https://doi.org/10.17576/jsm-2019-4809-26

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*, 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wiese, H., Allen, S.E.M., Keller, M., & Alexiadou, A. (2025). Introduction: Investigating the dynamics of language contact situations. In S.E.M. Allen, M. Keller, A. Alexiadou, & H. Wiese (Eds.), *Linguistic dynamics in heritage speakers: Insights from the RUEG group* (pp. 1–29). Language Science Press. https://doi.org/10.5281/zenodo.15775156

Xie, Z. (2018). The influence of second language (L2) proficiency on cognitive control among young adult unbalanced Chinese-English bilinguals. *Frontiers in Psychology*, *9*, 412. https://doi.org/10.3389/fpsyg.2018.00412

Yang, Y., Wu, S., Duñabeitia, J. A., Jiang, K., & Li, Y. (2021). The influence of L2 proficiency on bilinguals' creativity: The key role of adaptive emotion regulation strategies during the COVID-19 pandemic. *Frontiers in Psychology*, *12*, 695014. https://doi.org/10.3389/fpsyg.2021.695014

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin*, *124*, 283–307. https://doi.org/10.1037/0033-2909.124.3.283

Yu, B. (2013). Issues in bilingualism and heritage language maintenance: Perspectives of minority-language mothers of children with autism spectrum disorders. *American Journal of Speech-Language Pathology*, *22*, 10–24. https://doi.org/10.3389/fpsyg.2021.695014

Zeidan, J., Fombonne, E., Scorah, J., Ibrahim, A., Durkin, M. S., Saxena, S., Yusuf, A., Shih, A., & Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. *Autism Research*, *15*, 778–790. https://doi.org/10.1002/aur.2696

Zhang, Y., & Aryadoust, V. (2022). A systematic review of the validity of questionnaires in second language research. *Education Sciences*, *12*, 723. https://doi.org/10.3390/educsci12100723

Ziatabar Ahmadi, S. Z., Jalaie, S., & Ashayeri, H. (2015). Validity and reliability of published comprehensive theory of mind tests for normal preschool children: A systematic review. *Iranian Journal of Psychiatry*, *10*, 214–224.