Table S1

Databases and Search Terms Used

Databases	Search Terms
SCOPUS	1 bilingual OR multilingual OR "dual language" OR "foreign language learn" OR "english as a second language" OR "second language learn" OR "english language learn*" 2 "theory of mind" OR mentalization OR mentalizing OR "Role-Taking" OR "role taking" OR "perspective taking" OR "perspective-taking" OR "false-belief" OR "false belief"
PsycINFO (1806-Ovid)	1 bilingual*.ti,ab. 2 multilingual*.ti,ab. 3 dual language.ti,ab. 4 foreign language learn*.ti,ab. 5 english as a second language.ti,ab. 6 second language learn*.ti,ab. 7 english language learn*.ti,ab. 8 exp Multilingualism/ 9 exp "Theory of Mind"/ 10 exp Mentalization/ 11 theory of mind.ti,ab. 12 mentaliz*.ti,ab. 13 perspective taking.ti,ab. 14 role taking.ti,ab. 15 false belief.ti,ab. 16 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 17 9 or 10 or 11 or 12 or 13 or 14 or 15 18 16 and 17
Medline (Ovid) 1946	1 exp Bilingualism/ 2 *Multilingualism/ 3 bilingual*.ti,ab. 4 multilingual*.ti,ab. 5 dual language.ti,ab. 6 exp Foreign Language Learning/ 7 exp "Theory of Mind"/ 8 exp Mentalization/ 9 exp Role Taking/ 10 theory of mind.ti,ab. 11 mentaliz*.ti,ab. 12 perspective taking.ti,ab. 13 role taking.ti,ab. 14 second language learn*.ti,ab.

15 english language learn*.ti,ab.
16 exp False Beliefs/
17 false belief.ti,ab.
18 1 or 2 or 3 or 4 or 5 or 6 or 14 or 15
19 7 or 8 or 9 or 10 or 11 or 12 or 13 or 16 or 17
20 18 and 19

Table S2

Reasons for Excluded Papers

Reasons	Count
Retrieval issues	7
Not relevant after full read	17
Different population or developmental trajectories	
ASD (autism spectrum disorder)	2
Hearing impairment and sign language	7
No bilingual participants	1
Repeated publications	
Dissertations that were published, removed to avoid duplicating results	4
Studies used the same data and method as their previous study, only varied in analysis	1
Publications that repeat another study's result due to publishing in a different format (such as a chapter or under a different name)	5
Missing key constructs relevant to the review	
Intervention study	1
Meta-analysis	1
No methods section	7
No ToM	3
Grand Total	**56**

Reference	N	Bilingual Characterization						Relevant Study Objective(s)	Relevant Results
		Mono/Bilingual classification	Questionnaire?	LP	LH	LE	LC		
Gorrell et al. (1982)	40 Children 6 y.o. 20 English monolinguals 10 Viet-American bilinguals 10 Spanish-American bilinguals	5	0	0	0	0	0	To evaluate whether bilingual children show different spatial-role-taking abilities compared to monolinguals	• Main effect for bilingualism and culture, scoring higher than control monolingual groups specifically on the Block Design subtest of WISC-R • Viet-English scored higher than control and Spanish-English bilinguals • No results on the other spatial tasks
Gorrell (1987)	57 Children 5-6 y.o. 3 groups (no N given: monolingual English, monolingual Spanish, bilingual Spanish-English)	2	0	1-B	0	0	0	Building off Gorrell et al., (1982)'s work on spatial perspective taking differencers between bilinguals and monolinguals	• There are no differences between bilinguals and monolinguals • Age difference exist (regardless of bilingual or monolingual), where older children outperform younger children
Sperling (1990) Dissertation	48 Children 4-7 y.o. 24 Spanish-English bilinguals 24 English monolinguals	2	Home Interview Questionnaire (HIC; no source)	1-B	0	0	2, 3	To examine bilingualism and role-taking skills	• Generally, bilingual performed better than monolinguals on the Borke-Task and Chandler Task • There are several nuances depending on the task, which story, and type of descriptions
Jean-Louis (1999) Dissertation	83 Children 3-4 y.o. 28 English monolinguals 30 French monolinguals 25 French-English bilingual	1, 3, 4	1	1-B	0	1	1, 2, 3	To test which cognitive development in bilinguals differs: megalinguistc skills, ToM, or memory	• Significant age effect (older answer more correctly) in ToM tasks • No difference in ToM between monolinguals and bilinguals • No correlation between ToM and other constructs measured except negative correlation between 2 factors in the metalinguistic task (context explanation and change of object name)
Frank (2000) Dissertation	126 Children 2-3 y.o. 40 L1 English 13 L1 French 31 English-French bilinguals 32 French-English bilinguals	4	1	1-B	1	1	1, 2, 3	To investigate mutual exclusivity, relevant is its relation in bilingual's visual perspective taking	• In regard to just the perspective-taking results, there is no difference between the language groups, only an age effect where older performed better
Rodriguez. (2000) Dissertation	71 Adolescents ~11-13 y.o. 26 Spanish dominant monolingual 23 English dominant monolingual 22 balanced bilingual (see characterization)	1, 2	1	1-B	0	0	2, 3	To examine the relationships between language proficiency and perspective-taking/empathy in a bilingual population	• There is no significant difference between all three groups for empathy and perspective taking
Goetz, P. J. (2003)	104 Children 3-4 y.o. 32 English monolinguals 32 Mandarin monolinguals 40 Mandarin-English bilinguals	4	1	1-B	1	1	1, 2, 3	To determine whether linguistic knowledge (measured through PPVT) affect development of ToM	• No difference between the two groups of monolinguals • Bilinguals performed better in all ToM tasks but results vary depending on tasks
Berguno & Bowler. (2004)	197 Children 3-4 y.o. 140 "single language users" 57 "dual language users"	1	0	0	0	0	3	To examine the effects of knowledge/use of second language on children's understanding of FB and appearance-reality distinction	• Dual language children scored better than single language children on both tasks
Bialystok & Senman. (2004) Study 2 only	95 Children 4-5 y.o. 52 monolinguals 43 bilinguals (various languages)	1	0	1-L2	0	0	2, 3	To determine if there are developmental difference in A-R performance by examining a different language background (monolingual vs. bilingual)	• Only when controlling for vocabulary (PPVT-R), bilinguals scored better than monolinguals on reality questions • No difference regarding the appearance questions
Chan. (2005). Dissertation	60 Children 2-6 y.o. 29 English monolinguals 31 Chinese-English bilinguals	1, 2, 3	Bilingual Family Questionnaire (adapted from Bilingual Family Telephone Questionnaire (Paradis, Crago, Genesee, & Rice, 2003))	1-B	1	1	1, 2, 3	To determine whether bilinguals have more advanced cognitive development and ToM than monolinguals	• Bilingual status is a predictor of ToM and other cognitive tasks • Simialrly main effect with bilingual status on ToM only when controlling for verbal and non-verbal abilities

Study							Purpose	Relevant Findings
Kobayashi et al. (2006)	32 Adults ~28 16 Japanese-English bilinguals 16 American English monolinguals	0	0	0	1	0	0	To explore if there are cultural and linguistic variations in neural regions in adult ToM by examining bilinguals vs. monolinguals	Note specific neural regions are not noted down here. Only general findings relevant to bilingualism are indicated. • Cultural difference noted due to different regions activated during ToM tasks for monolinguals vs. bilinguals • Different activations between L1 version of the task vs. L2, suggesting bilinguals process the ToM task differently depending on language. • Note that for the tasks, no group differences were found, only neural correlates
Pelletier. (2006)	Study 1: 79 Children 9.5 y.o 36 English monolinguals 43 L1 Portuguese attending English school Study 2: 228 Children 4-5 y.o. 73 L1 English All bilinguals with L2 English: 54 L1 Cantonese 42 L1 Tagalog 59 L1 Ukrainian	1	0	1-B	0	0	3	To investiage how ToM, metacognitive language, phonlogical processing and reading comprehesnion are related (study 1) with addition of cognitive factors such as short term memory (study 2)	• Relevant results indicate L1 English scored higher on vocabulary and ToM • ToM performance is correlated with fables task, hence comprehension and it predicted vocabulary development (study 2 only)
Kobayashi et al. (2007)	24 Children 8-11 y.o. 12 English monolinguals 12 Japanese-English bilinguals	0	0	1-B	1	0	0	As an extension to Kobayashi et al., 2006 (see above), to examine if there are cultural and linguistic dependent or independent neural correlates in children's ToM	Note that neural regions are not noted down. Only general findings: • vmPFC as main region that activates during ToM tasks for all groups • There are certain differences in activation between
Kobayashi et al. (2008)	16 Adult Japanese-English bilinguals (~29 y.o.) 12 Children Japanese-English bilinguals (~10 y.o.)	0	0	1-B	1	0	0	To explore linguistic effect on neurl areas for ToM between late bilingual adults and early bilngual children	Note only general findings are reported • mPFC is shown to activated during ToM tasks • Adults activated more dorsal mPFC for L1ToM, and ventral for L2 ToM and is more language dependent than children
Kovacs. (2009)	64 Children 2-4 y.o. 32 Romanian monolinguals 32 Romanian-Hungarian bilinguals	4	1	1	0	1	2	To investigate how growing up as "crib" bilingual would affect performance on false-belief tasks	• Bilinguals performed better than monolinguals on both ToM tasks • Attributed to the fact the task could be showing an inhibitory control due to code-switching and thus performed better in general
Kyuchukov & DeVilliers. (2009)	Study 1: 30 Children 3-5 y.o. All Romani-Bulgarian bilinguals Study 2: 120 Children 3-5 y.o. 60 Bulgarian monlinguals 60 Romani-Bulgarian bilinguals	0	0	0	0	0	1	To evaluate if ToM mastery is different in bilingual Roma children vs. monlinguals	• Study 1 indicated Romani children perform ToM tasks better in L1 than L2 • However, study 2 showed no difference bewteen L1 and L2, only as age increased did ToM performance increase for both language groups
Cheung et al. (2010)	121 Children 3-4 y.o. 59 second language learners 62 bilinguals (see characterization)	1, 4	1	1-L2	1	0	1, 2, 3	To examine a sociolinguistic awareness and its relation to false belief in bilingual children	• After adjusted for age, nonverbal intelligence, SES and vocabulary bilinguals scored higher than second language learners • Sociolinguistic awareness is a preditor of false belief
Farhadian et al. (2010)	163 Children 4 y.o. 65 Persian monolinguals 98 Kurdish-Persian bilinguals	0	0	1	0	0	0	To see if there is difference between bilingual and monolingual ToM and if ToM is related to verbal abilities and age	• Older children performed better • When controlling for age, verbal abilities predicted FB scores • Bilinguals scored significantly better than monolinguals on FB tasks • Regression model indicate language status (bilingual or monolingual) predicted ToM above and beyond age, and verbal abilities

Study	Sample	C1	C2	C3	C4	C5	C6	Purpose	Findings
Tare & Gelman. (2010)	28 Children ~2-5 y.o. All Marathi-English bilinguals	2	1	1-B	1	0	1, 2	To examine bilingual children's pragmatic skills in conversation through free play and its connection with ToM and metalinguisic awareness	• Older children performed better on ToM tasks • When controlling for age, ToM is associated with increase responsiveness to switching languages • Metalinguistic awareness is correlated with children being more likely to switch language • Childen switched language, but spoke the appropriate language majority of the time (i.e., English with English experimenter)
Rubio-Fernandez & Gluksberg. (2012)	46 Adults ~19y.o. 23 English monolinguals 23 bilinguals	4	Language History Questionnaire (Gullberg & Indefrey, 2003)	2	1	1	0	To test if bilingual adults show similar difference in FB reasoning compared to monolingual adults	• Adults did have egocentric biases with FB performance, based on eye-tracking • Sally-Anne (EF) is correlated with FB for both monolinguals and bilinguals • Bilinguals were less susceptible to egocentric biases than monolinguals due to faster fixation on the correct target.
Greenberg et al. (2013)	82 Children 8 y.o. 45 English monolinguals 37 bilinguals	4	1	1-L2	1	1	2	To determine if executive control in bilingual children is related to complex perspective taking task and ToM	• Significant effect for language group where bilinguals were more accurate than monolinguals • Regression model showed PPVT, verbal intelligence, and language group were significant predictor of perspective taking
Han & Lee. (2013)	133 Children 4-5 y.o. 60 monolinguals 73 bilinguals	1, 2	1	1-B	0	0	2, 3	To examine the cognitve and affective perspective-taking abilities in balanced bilinguals in South Korea	• Older children performed better than younger on the cognitive task. There is no difference between monolinguals and bilinguals • On the affective task, bilinguals performed better than monolingual children, but there was no age difference
Pearson. (2013) Dissertation, only study 4	68 Children ~3y.o. 40 English monolinguals 28 English-Spanish bilinguals	0	0	1-L2	0	0	1	To determine if metalinguistic abilities or EF is related to bilingual ToM development	• No significant difference between monolingual and bilingaul FB, though both group scored low on this FB task
Batres. (2014) Dissertation	96 Adults ~20 y.o. 48 English Monolinguals 48 English-Spanish Bilinguals	1, 4	Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld & Kaushanskaya, 2007)	1-B	1	1	2, 3	To investigate the relationship between perspective-taking abilities and cognitive control	• Bilinguals performed slower than monolinguals when switching to new label in the tangram-matching task • In general, bilinguals performed worse
Nguyen & Astington. (2014)	72 Children 3-5 y.o. 24 English monolinguals 24 French monolinguals 24 English-French bilinguals	2, 4	1	1-B	1	1	0	To control for SES and examine whether inhibition of working memory is the cause of the bilingual difference in FB tasks	• Bilingual difference only when controlling for age and verbal abilities (verbal abilities are lower in bilinguals) • Mediation model was drafted and indicated an indirect effect of BWS (working memory measurement) on FB
Ryskin et al. (2014)	Study 1: 64 Adults 18-26 y.o. 31 English monolinguals 33 bilinguals (various languages) Study 2: 41 Adults 18-23 y.o. 21 English monolinguals 20 bilinguals Study 3: 41 Adults ~20 y.o. 22 Monolinguals 19 Bilinguals	4	1	1-L2	1	1	0	To look at the effects of bilingualism on visuo-spatial perspective taking in adults	• Bilinguals showed no difference EF nor difference in the perspective-taking conditions • Instead bilinguals performed worse on the no-perspective taking condition
Fan et al. (2015)	72 Children 4-7 y.o. 24 English monolinguals (little exposure to a second language) 24 exposure group (regular but limited exposure to a second language) 24 bilinguals (regular exposure to English and another language)	3	0	1-L2	1	1	2, 3	To evaluate if children exposed in different language environment would be able to understand another's perspective	• Language group had main effect, where exposure and bilingual group were more correct than the monolingual group • Bilingual and exposure group also recovered (eye gaze) faster than monolinguals • Congitive factors were not predictors of perspective-taking

Study	Demographics							Purpose	Findings
Mante-Estacio & Bernardo. (2015)	Study 1: 76 Adolescents 16-19 y.o. Study 2: 100 Adolescents 16-18 y.o. All Filipino-English bilinguals	2	0	0	0	0	0	To investigate whether illusory transparency (reader incorrectly assuming story character's perspective) is different in bilinguals (study 2 varies modalities)	• Study 1 showed limited non-significant results • All showed illusory transparency, that they mistake the perspective, but there is an interaction effect between language and the illusory effect with more errors in the English text
Yow & Markman. (2015)	32 Children 3 y.o. 16 monolinguals 16 bilinguals (various languages)	4	1	1-L2	1	1	2	To test whether bilingual are able to understand speaker cues and communicative intent	• No difference between monolingual and bilingual on vocabulary, STM, and inhibition • Bilinguals outperformed monolinguals when it came to picking up multiple cues and successfully interpreting the "where" condition
Banasik and Podsiadło. (2016)	31 Children ~5-6 y.o. All Polish-English bilinguals	1, 3	0	1-L1	0	0	3	To study irony comprehension in bilingual children and its relation with ToM	• No ToM data interpretation done by the authors as the focus is irony • Irony and ToM are correlated
Cox et al. (2016)	90 Adults ~73 y.o. 64 monolinguals 26 bilinguals (language varies)	4	1	1	1	1	0	To assess the directionality of bilingual's executive function and social cognitive differences. Focus here will be the ToM measurement	• Regarding the faux-pas test, bilinguals did not perform better than monolinguals
Gordon. (2016)	52 Children 3-6 y.o. 26 English monolinguals 26 Spanish-English bilinguals	1, 3	1	1-B	1	1	2, 3	To investiage how language proficiency is related to ToM performance among monolinguals and bilinguals	• Only diverse desires (bilinguals scored higher) and explicit false belief (monolinguals scored higher) versions of the task showed significance • English proficiency predicted perfomance of tasks for monolinguals, not bilinguals, but instead, bilinguals needed high proficiency in both languages to predict FB scores
Weimer & Gasquoine (2016)	102 Children ~3-7 y.o. 26 English dominant 23 balanced bilingual 53 Spanish dominant	1, 2	0	1-B	0	0	0	To study the predictors of emotional understanding and belief understanding (authors' definition of ToM)	• Age and emotional understanding predicted belief reasoning • Vocabulary is positively correlated with belief reasoning and emotional reasoning
Dahlgren et al. (2017)	28 Children 3-5 y.o. 14 Swedish monolinguals 14 Swedish-Slovanian bilinguals	0	0	1-L1	0	0	1	To compare ToM and EF between bilingual and monolingual children	• Bilingual performed better on EF task for attention, inhibition (controlling for vocabulary) than monolinguals • However, bilinguals did not outperform monolinguals on ToM tasks
Hsin & Snow. (2017)	82 Children ~9-11 y.o. 41 English monolinguals 41 "language-minority" student	1, 3	0	0	0	0	0	To examine perpective-taking in argumentative essays among language minority students versus monolingual students	• Language minority students show more perspective-taking in their writing compared to English-only students
Diaz & Farrar. (2018a).	65 Children 3-5 y.o. 33 English monolinguals 32 Spanish-English bilingual	3	1	1-B	1	1	2, 3	To examine the relationship between EF and language proficiency on FB	• Only after controlling for language abilities do bilinguals performed better than monolinguals on FB • EF does not have an effect on bilingual's FB performance but instead had an effect for monolingual (inhibition through Bear/Dragon task)
Diaz & Farrar. (2018b).	78 Children ~4 y.o. 38 English monolinguals 40 Spanish-English bilinguals	3	1	1-B	1	1	2, 3	To use a longitudinal design and investigate the influences of EF, language ability, and metalinguistic awareness on FB	• As monolinguals outperformed bilinguals in vocabulary, when controlling for it, bilinguals performed better on FB tasks • Metalinguistic abilities at time 1 predicted bilingual's FB performance at time 2 (1 year later) • For monolinguals, inhibition and language were predictive of FB
Grover. (2019)	302 Children 3-5 y.o. All bilinguals with Norwegian as L2	3	1	1-B	0	0	1, 2, 3	To explore what predicts narrative perspective-taking in children	• After controlling for age and L2 vocabulary, narrative skills predict perspective taking • Bilingual language use (using L2 at home) also predicted perspective-taking
Lorge & Katsos. (2019)	40 Adults 20-35 y.o. 20 English monolinguals 20 bilinguals (various languages)	4	1	0	1	1	1	To examine how speaker characteristics can affect communication skills in listener-adapted speech	• Results focused on linguistic features, noting that bilinguals did adapt their speech to the child and the "foreigner" conditions using more repetition, pitch change, speed rate etc. • Personality (specifically conscientiousness) correlated with perspetive-taking

Study	Demographics		Language Questionnaire						Aim	Findings
Raisa et al. (2019)	60 Children 3-8 y.o. All Kannada-English bilinguals	1, 4	1	1-B	1	1	1		To explore how language influences higher-order ToM in Indian bilingual children	• Authors noted a developmental trend, with younger children unable to answer the 3rd level ToM questions, attributed to their language
Buac & Kaushanskaya (2020)	115 Children 5-10 y.o. 44 English monolinguals 44 English-Spanish simultaneous bilinguals 27 English-Spanish bilinguals through dual language immersion programs	4	Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld & Kaushanskaya, 2007)	1-B	1	1	2, 3		To examine whether language skills and EF predict ToM	• Language skills predicted ToM for simultaneous bilinguals • Working memory predicted ToM for monolinguals • Inhibition and shifting predicted ToM for English L1 bilinguals
Dicataldo & Roch. (2020)	115 Children 3-6 y.o. All some level of bilingual exposure	1, 3	1	1-L2	0	1	2, 3		Note broad study, only ToM construct is indicated here, specifically to investigate variation in the bilingual exposure on development.	• Specifically related to ToM, SES and more exposure to languages resulted in higher ToM scores • ToM isn't correlated with inhibition and attention in this study
Barber et al. (2021)	84 Children ~8-10 y.o. All "emergent bilinguals"	3	0	1-B	0	0	2, 3		To investigate the relationship between ToM with reading and listening comprehension in bilinguals	• Bilingual ToM predicted listening and reading comprehension after controlling for grade, word ID, and vocabulary
Kim et al. (2021)	317 Children ~6-8 y.o. All Spanish-English dual language learners (69% in dual language programs, 31% in English immersion programs)	1	0	1-B	0	0	3		To expand on L1-L2 writing skills and its connection with higher level cognitive abilities such as inference making, perspective-taking, and comprehension monitoring	• Higher congitive skills such as ToM is related and predicted writing • Results takes sex, SES (poverty status), grade, school, English learner status, and biliterate status into consideration of the final model
Navarro. (2021) Dissertation, only study 2	154 Adults ~38 y.o. 92 monolinguals 62 bilinguals	4	Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007), & Language and Social Background Questionnaire and the Bilingual Switching Questionnaire (Rodriguez-Fornells et al., 2012)	1-B	1	1	1, 2, 3		To compare ToM performance in bilingual and monolingual adults and whether metalinguistic awarness and EF	• Bilnguals performed worse than monolinguals a group level, but individually, bilingualism did predict ToM performance • Note a large number of bilinguals underperform on the verbal task • Director task is predicted by Simon's task and metalinguistic tasks for both language groups
Navarro & Conway. (2021)	78 Adults ~27 y.o. 41 English monolinguals 37 bilinguals	4	Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld & Kaushanskaya, 2007)	2	1	1	1, 2, 3		To ask if adult bilinguals also show improved ToM performance compared to monolinguals	• No different in reaction time but bilingual were more accurate than monolingual on the Director condition • Using the indicated culture index, multicultured participants outperformed monoculture in the experimental Director condition as well
Singh et al. (2021)	55 Children 3-6 y.o. All Mandarin-English bilinguals	4	Language Exposure Questionnaire (Bosch & Sebastián-Gallés, 1997)	1-B	0	1	1, 2, 3		To examine how cognitive difference such as ToM in bilinguals predict racial baises	• Note majority did not perform at ceiling on the FB task • Perspective-taking (FB task) is not related to racial baises
Stegall-Rodriguez et al. (2021)	147 Children 4-9 y.o. 58 monolinguals 62 language dominant bilinguals (see bilingual breakdown) 15 balanced bilinguals	2, 3	1	1-B	0	0	0		To examine low SES biligual and their ToM, ambiguous figures and inhibitory control	• Controlling for age and inhibition shows a relationship between ToM and ambiguous figures • No significant findings regarding ToM and relation with inhibition or other variables
Sudo & Matsui. (2021)	50 Children ~5-6 y.o. 25 Japanese monolinguals 25 Brazilian dual language learners	1, 5	0	1-B	1	0	1, 2, 3		To examine if dual language learners' school readiness in areas of language, EF, and ToM in a Brazillian-Japanese context	• While there is no difference in inhibition, monolinguals scored higher than DLLs on verbal FB tasks, associated with a lower language ability of DLLs
Tarighat & Krott. (2021)	216 Adults ~33 y.o. 108 Persian monolinguals 108 bilinguals (Persian is L1 for a small group)	4	1	2	1	1	1		To examine if there is a gender difference in bilingual performance on perspective-taking	• Female scored higher than male on the IRI • However, only bilingual men were better than monolingual men

Tiv et al. (2021)	66 Adults 18-31 y.o. All bilinguals but vary in L1 English and L2 English	4	1	2	1	1	1	To ask whether bilinguals (and language diversity measured through language entropy) differ in their mentalizing abilities, focusing on reading English in L1 and L2.	• While all readers judge logical better than mental state condition, greater language diversity is related to more mentalizing and not the logical condition in the statistical model • Possible "over-mentalization" in L2 readers as they apply it to all conditions
Wimmer et al. (2021)	67 Children 3-5 y.o. 34 English monolinguals 33 bilinugals (L1 = English, L2 = various languages)	4	1	1-L2	1	0	2, 3	To evaluate bilingual performance in ambiguous figures and the relationship with inhibition	• In regarding to ToM, no different in FB task and Droodle between monolinguals and bilinguals • Bilinguals performed better on the reversal indicating better inhibition, but isn't predicted by FB/Droodle
Gasiorek et al. (2022)	197 Adolescents 16-19 y.o. 97 Swedish monolinguals 100 Swedish-Finnish bilinguals	1, 3	0	2	0	0	2	To examine whether home language and perspective taking predict language choice in adolescents	• Adolescents from bilingual households are more aware of other language preferences

Reference	ToM Task Name and Reference	Task Description and Modifications	Administration		Answer		Task Language	Additional Considerations			
			VT or NT	Testing format	VR or NR	Response format		EF	MA	SES	Cog or Aff
Gorrell et al. (1982)	Spatial orientation (resembles Hegarty & Waller, 2004)	Various objects (colours, blocks, dolls, pictures of human faces etc.) are arranged in certain positions. Children are asked to select a response that matches the perspective of another child.	2	Pictures, objects	2	MC (pictures)	L2* (Unclear)	0	0	1	C
Gorrell (1987)	Spatial orientation	Used props and have children select 1 out of 4 choices to represent the view of someone else. One choice is the egocentric view. Children were able to move blocks freely.	2	Pictures, objects	2	MC (pictures)	L1	0	0	0	C
Sperling (1990) Dissertation	• Borke's Interpersonal Awareness Task (Borke, 1971) • Chandler's Bystander Cartoon Sequence (Chandler, 1973)	In Borke's task, pictures and stories are read and shown to participants, then identify emotions on drawn faces based on a story read to them. In Chandler's task, children describe pictures and then retell them from the perspective of a bystander.	3	Pictures, stories	1	Verbal response, MC (pictures), story retell	C	0	0	1	D
Jean-Louis (1999) Dissertation	False-belief Task (Wimmer & Perner, 1983) > Change in location	- Standard verbal story with a change in location, delivered with 3 questions (belief, reality, memory) to check understanding.	1	Stories, verbal questions	1	Verbal response	B	1	1	1	C
Frank (2000) Dissertation	Visual perspective-taking task (loosely based on Flavell et al., 1968)	Two perspective-taking tasks using pictures, similar to a visual-spatial task require taking the visual perspective of the experimenter. Vary in verbal vs. nonverbal. Requires rotation of a physical object to the correct perspective	2	Pictures, objects	3	Verbal response, object rotation	L1	0	0	1	C
Rodriguez. (2000) Dissertation	Scenario response perspective task (based on Bengtsson & Johnson, 1992)	Participants are read a scenario usually involving affect and are asked to imagine themselves in in the hypothetical situations. They are then asked "What do you think about that? What do you think when you see/hear that?"	1	Stories	1	Open-ended	B	0	0	1	A
Goetz, P. J. (2003)	• Appearance-reality test (Flavell et al., 1983) • Level 2 perspective-taking task (Flavell et al., 1981) • Unexpected contents false-belief task (Hogrefe et al., 1986) • Unexpected transfer false-belief task (Wimmer & Perner, 1983)	Standard delivery of classic appearance-reality and false-belief tasks using objects and story vignettes respectively	3	Objects, stories, verbal questions	1	Verbal response	B	0	0	1*	C
Berguno & Bowler. (2004)	Appearance-reality task adapted from Flavell et al. 1983	Fish/pen version, asked 3 questions about a trick object (reality, appearance, false belief (egocentric))	1	Objects, verbal questions	1	Verbal response	L2	1	0	0	C
Bialystok & Senman. (2004) Study 2 only	Appearance-reality task (Flavell et al. 1983; with modifications based on Gopnik & Astington, 1988)	Four duo objects. Standard AR procedure. Asked "can you tell me what this is?"	1	Objects, verbal questions	1	Verbal response	L2* (Unclear)	1	0	1*	C
Chan. (2005). Dissertation	• Box/Basket Task • Desk/Cupboard Task (above two are both nexpected content task based on Jenkins & Astington, 1996) • Book Task • Crayon Box Task (Change in location)	Followed an adaptation fo the classic tasks based on Jenkins & Astington (1996). Involved objects and asking about false-beliefs questions.	1	Objects, stories, verbal questions	1	Verbal response	L2	1	1	1	C
Kobayashi et al. (2006)	Second-order ToM task (Perner & Wimmer, 1985)	Due to fMRI, this was delivered in block design, the participants were reading the vignettes, and a baseline using non-ToM tasks were established	1	Reading, stories	1	MC (text)	B	0	0	0	C
Pelletier. (2006)	A second order task (modified to make more difficult for older children) (based off Astington, Pelletier, & Homer, 2002)	Ask the participant to respond to a story and must explain their reasoning	1	Stories	1	Open-ended	B	1	1	0	C
Kobayashi et al. (2007)	Identical design to Kobayashi et al., 2006 with the addition of a cartoon task	An additional task with a cartoon story presented in the block design	3	Reading, pictures	3	MC (text, and pictures)	B	0	0	0	C
Kobayashi et al. (2008)	Identical design to Kobayashi et al., 2006	See above	1	Reading	1	MC (text)	B	0	0	1*	C
Kovacs. (2009)	• False-belief Task (Wimmer & Perner, 1983) • Modified ToM/False-belief Task to simulate language switching.	The modified task using pictures and dolls. Requires both languages to understand the need for language switching.	1	Stories, pictures	1	Verbal response, language switching	B	0	0	1	C

Study	Task	Description		Stimuli		Response	Level					
Kyuchukov & DeVilliers. (2009)	• Unexpected content (Wellman, Cross & Watson, 2003) • Unseen dispalcement	Standard 3 questions delivered with the regular version of the tasks. Used chocolates/biscuit for the unexpected contents.	1	Objects, stories	1	Verbal response	B	0	0	1		C
Cheung et al. (2010)	• Unexpected content • Unseen displacement • Sociolinguistic awareness task	The sociolinguistic awareness task required children to pick up language cues and confusion by switching languages that are appropriate	1	Objects, stories	1	Verbal response, language switching	L1	0	0	1		C
Farhadian et al. (2010)	• Sally and Ann • Red/Blue Box (change of location; Baron-Cohen et al., 1985) • Crayon Box/Sticker (change of content; Gopnik & Astington, 1988)	Standard procedure using stories followed by 3 questions (false belief and memory)	1	Objects, stories	1	Verbal response	L2	1	0	1		C
Tare & Gelman. (2010)	• 3 separate tasks from Wellman & Liu (2004): diverse desire, diverse belief, and knowledge access	Also included a language switching task during free play where the experimenter may speak one of two languages	1	Objects, stories	1	Verbal response	B	0	1	0		C
Rubio-Fernandez & Gluksberg. (2012)	Sally-Anne task using both false-belief and true-belief condition	Alongside the FB tasks (standard procedure), RT and eye-tracking were done.	3	Computer video	3	Verbal response, eye tracking	L2	1	0	0		C
Greenberg et al. (2013)	Computerized visual perspective-taking task	Computer task of image of an owl, requires rotation to successfully grasp the perspective of the fictional character.	2	Computer object	2	MC	L2* (Unclear)	0	0	0		C
Han & Lee. (2013)	• Cognitive perspective-taking task (Flavell, 1968, modified by Kurdek & Rodgon, 1975) • Affective perspective task (Borke, 1971, modified by Kurdek & Rodgon, 1975)	Children shown set of pictures, then some pictures were removed in the cognitive task that requires the participant to retell the story to someone who did not see the pictures. Affective task also used pictures but is focused on responding to character emotions that may be appropriate or not to the situation	3	Pictures and stories	1	Story retell	C	0	0	1		D
Pearson. (2013) Dissertation, only study 4	Sally and Ann variation using puppets (Peterson & Siegal, 1995, modified by Doherty, 2000)	Entirely verbal instructions with. Task takes a deceptive motive turn in the puppet condition but generally followed the same structure. Author's goal was to make it more robust.	1	Objects, stories, verbal questions	1	Verbal response	L2	0	1	0		C
Batres. (2014) Dissertation	Tangram-Matching task (Metzing & Brenna, 2003) developed to be similar to ambiguous objects where participants needed to switch their labelling based on partners	Matching geometric figures (tangrams). Children arranged the tangrams in specific orders. Objects varied in fake and real names from study 1 to 2 and required shift in language.	1	Computer object	1	Language switching, naming	B	1	0	0		C
Nguyen & Astington. (2014)	• Change in location task (Wimmer & Perner, 1983) • Unexpected contents task (Gopnik & Astington, 1988; Perner et al., 1987)	Standard delivery and procedure.	3	Computer video	1	Open-ended, verbal explanation	B	1	0	1		C
Ryskin et al. (2014)	• Map orientation task • Grid based task (Brown-Schmidt et al., 2008)	A complex set of visuospatial tasks that requires following verbal instructions. Perspective-taking version had an opposite or reversed map that required participants to reorient their perspective. The linguistic stimuli varied where the instruction much be the opposite perspective.	3	Comupter, stories, directions, pictures	1	Drawing, eye tracking, MC (pictures)	L2* (Unclear)	1	0	1		C
Fan et al. (2015)	Director's task (similar to Wu & Keysar, 2007)	Director tasks performed in person with a 4x4 grid of objects. Eye gaze were recorded to evaluate recovery and switching	1	Directions	3	Verbal response, MC (pictures), eye-tracking	L1	1	0	1		C
Mante-Estacio & Bernardo. (2015)	Illusory transparency effect task (Keysar, 1994)	Using several stories with ambiguous scenarios. Prompted with "How do you think the character-x feel about character-y?" on a Likert scale. The tasks vary in the type of information the reader may know (positive or negative)	1	Reading	1	MC (Text)	B	0	0	0		A
Yow & Markman. (2015)	Novel Object design with "there/where" manipulation (based on Nurmsoo, E., & Bloom, P., 2008)	Uses objects that are not seen by the experimenter. Manipulated a comment as "there it is!" versus "where is it?" to indicate the experimenter's perspective and awareness of the object. To succeed, children need to pick up eye gaze, context, and semantics	1	Objects, verbal questions	1	Verbal response, naming	L1	1	0	1		C
Banasik and Podsiadło. (2016)	Reflection on Thinking Test (TRM; Białecka-Pikul, 2012)	Uses stories that allowed quantitative data and a qualitative "why" answer. The task involves a series of different assessments, on visual perspective, intentions, false-belief and so on. Certain design is similar to other FB tasks with a story vignette, but remade into a Polish version.	3	Pictures, stories	1	Open-ended	L1*	0	0	0		C

Study	Task	Description		Input		Response					
Cox et al. (2016)	Faux Pas test (Stone et al., 1998; Gregory et al., 2002)	Standard faux paus, asking participants to identify something that is awkward (a faux pas) in a social context. Self paced task, must read each story	1	Reading	1	MC (Text)	L1	1	0	1	B
Gordon. (2016)	All 7 tasks outlined in Wellman & Liu, 2004 • Diverse desires • Diverse beliefs • Knowledge access • Contents false belief • Explicit false belief • Belief-emotion • Real-apparent emotion	Same procedure as original author. Assessed twice, 2nd version of the task was altered slightly to avoid chance responses	1	Objects, stories, verbal questions	1	Verbal response	L2	0	0	1	D
Weimer & Gasquoine (2016)	Adapted from Fabricius et al., 2010 version of the task • Unexpected content task • Change of location task • True belief unexpected contents Test of Emotion Comprehension (TEC; Pons et al., 2004) with addition of a belief-based emotion task from Weimer et al., 2012	Verbal delivery with objects. Questions are asked like "what do you think is inside the box?" etc. TEC asked participants to recognize emotions but also whether they understand emotions are based on desires, and people have hidden emotions.	1	Objects	1	Verbal response	B	0	0	1	D
Dahlgren et al. (2017)	• Eva and Anne (adapted from Sally-Anne, Baron-Cohen et al., 1985) • Kiki and the cat (Lewis, 1994) • Thought picture (Woolfe et alk., 2002) • Hide the fruit (Vinden, 1999)	Unclear how these were delivered, assuming it's similar to the procedure based on the original papers	3	Pictures, stories	3	Verbal response, MC	L2	1	0	1	C
Hsin & Snow. (2017)	• A writing task, Social Perspective-Taking Acts Measured (SPTAM) (Kim, LaRusso, Hsin, Selman, & Snow, 2016) that requires students to give advice to someone through short written responses	Asked students to give advice to someone through a short written response (in the form of essays)	1	Computer, reading	1	Written	L2	0	0	0	C
Diaz & Farrar. (2018a).	• Unexpected content task (Perner, Leekman, & Wimmer, 1987) • Unexpected location task (Wimmer & Perner, 1983) • Object disappearance task (Wellman et al., 2001) • Apperance-reality task with two versions, one to identify the object, second to identiy the properties of the object	Used props and has children select 1 out of 4 choices to represent the view of someone else. One choice was the egocentric view. AR tasks used objects (sponge, fish)	1	Objects, stories, verbal questions	1	Verbal response	C	1	0	1	C
Diaz & Farrar. (2018b).	• Unexpected location task (Wimmer & Perner, 1983) • Unexpected content task • Appearance–reality task	Standard delivery based on the original task. Verbal questions for each one.	1	Objects, stories, verbal questions	1	Verbal response	C	1	1	1	C
Grover. (2019)	A narrative task using a picture book called Hug by Alborough, 2012	Asks children to retell the story from the perspective of two different characters in the story: one requires just a narrative production, other requires the perspective of a different character who may not know all the information	1	Reading	1	Story retell	L2	0	0	1	B
Lorge & Katsos. (2019)	• A language adapted speech production task • Computerized version of the Director's task (Dumontheil, Apperly, & Blakemore, 2010; Keysar, Barr, Balin, & Brauner, 2000).	The production task gives a recipe to see if participants take the interlocuter into perspective and adjust their production. The task varied in two conditions, child-directed and foreigner-directed speech. Participants were asked to consider the foreign accent of the interlocutor.	1	Computer	1	Speech production, MC (pictures)	D	1	0	0	C
Raisa et al. (2019)	Story task (based on Liddel & Nettle, 2006)	Used 2 stories (one in Kannada, one in English) with a set of questions ranging from first-order to third-order ToM.	1	Stories	1	Verbal response	B	0	0	0	C
Buac & Kaushanskaya. (2020)	• First-order ToM using Sally-Anne task • Second-order is similar, but asks a second-order question which require taking on the perspective of the characters in the scenes.	Also included a visual and text.	3	Stories, pictures	1	Verbal response	D	1	0	1	C

Study	Task	Notes		Materials		Response	Code				
Dicataldo & Roch. (2020)	False-belief task (loosely based on Gopnik, & Astington, 1988)	In two trials, one is the egocentric answer, other is the perspective taking of a 2nd person. We considered this as two separate tasks.	1	Objects, verbal questions	1	Verbal response	L2	1	0	1	C
Barber et al. (2021)	Subtest from NEPSY-II (Korkman, Kirk, & Kemp, 2007)	Included situations to listen to, some with pictures and answering questions regarding different points of view, facial emotions	1	Standardized assessment (listening, pictures, question)	1	MC (Text)	L2	0	0	1	B
Kim et al. (2021)	• Inference task of the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999) • Theory of Mind Inventory-2 (ToMI-2; Hutchins et al., 2012) to assess ToM in both languages	Spanish version of the inference-making task was developed by the research team. Normed tasks but also developed their own versions that are "easier"	1	Standardized assessment (Audio recordings), pictures, questionnaires	3	Open-ended, MC (pictures)	B	0	0	1	B
Navarro. (2021) Dissertation, only study 2	Director Task (Dumontheil, Apperly, and Blakemore, 2010; Legg, E. W., Olivier, L., Samuel, S., Lurz, R., & Clayton, N. S., 2017)	Automated version, computerized with 2 trial conditions, identical to Navarro & Conway, 2021 (see below)	1	Computer, directions	2	MC (pictures)	D	1	1	0	C
Navarro & Conway. (2021)	Director's task (Dumontheil et al., 2010)	Both a director and no-director condition (where they do not need to keep the perspective of the director in mind). Included 2 trial conditions: experimental with competing objects, and control where there is only one possible object to consider	1	Computer, directions	2	MC (pictures)	D	0	0	0	C
Singh et al. (2021)	Sally-Anne Task (Baron-Cohen, Leslie, & Frith, 1985)	Standard delivery, names were changed to Billy and James	1	Objects, stories	1	Verbal response	D	1	0	1	C
Stegall-Rodriguez et al. (2021)	Unexpected contents task (also true-belief version)	True-belief version of the tasks does not require the perspective of another person and is not considered in this review.	1	Objects	1	Verbal response, open-ended	D	1	0	1	C
Sudo & Matsui. (2021)	Unexpected transfer (Wimmer & Perner, 1983)	Additional two trials (of the same task) for bilinguals that required little verbal responses (note however, this was not administered to monolingual group)	3	Stories, video	3	Verbal response	L1	1	1	0	C
Tarighat & Krott. (2021)	Interpersonal Reactivity Index (IRI; Davis, 1983)	IRI dates back to 1983. Authors specifically looked at the perspective taking subscale of the IRI. There are 7 questions that considers own perspective and others. Note the consideration of own perspective.	1	Questionnaire	2	MC (text)	D	0	0	1	C
Tiv et al. (2021)	Task was original with some adaptation from Ferstl & von Cramon, 2002; Nadig & Ozonoff, 2007; Lavoie, Vistoli, Sutliff, Jackson & Achim, 2016.	138 sentence pairs (context and action). Context varies in 3 conditions: logical inference, mental state inference, and incoherent. The mental state inference context condition requires participants to infer possible reasons for the action	1	Computer, reading	2	MC (text)	D	0	0	1	B
Wimmer et al. (2021)	• FB task (similar to Wimmer & Perner, 1983) • Droodle task (possible interpretation from Chandler & Helm, 1984)	For the droodle task, children were shown a portion of the image. After the whole image was shown. The perspective question asked if someone saw the zoomed-in image, what would they see?	1	Objects, stories	1	Verbal response	L2	1	0	0	C
Gasiorek et al. (2022)	• Situational perspective-taking task • Interpersonal Reactivity Index (IRI; Davis, 1983)	Assesses whether they consider the needs/preference of others when choosing a language. IRI matches description above (see Tarighat & Krott, 2021).	1	Reading, questionnaire	1	MC (text)	L2* (Unclear)	0	0	0	B