Supplementary Materials to:

A Systematic Review and Meta-Analysis of Language and Cognition in the Developing Bilingual

Brain: From Infancy to Adolescence

S1. Search strategies adapted for each database

Database Search Concepts and Keywords

Concepts	Keywords
(1) Method	'NIRS' OR 'MRI' OR 'MEG' OR 'TMS'
(2) Child	('Child' OR 'Infant' OR 'Adolescent' OR 'Teen*' OR 'Kid*' OR 'Preschool' OR 'School Age' OR 'Youth' OR 'Newborn*' OR 'Elementary School' OR 'Middle School' OR 'Kindergar*')
(3) Bilingual	'Bilingual*' OR 'Multilingual*' OR 'L2' OR 'Second/Dual/Home/Heritage/Immigrant Language'

Note. * truncation symbol

Searches combined with AND: 1 AND 2 AND 3

OVID Medline

Concepts	Keywords	Medical Subject Heading (MeSH)
(1) Method	fNIRS or NIRS).tw,kf. OR (fMRI or MRI).tw,kf. OR MEG.tw,kf. OR TMS.tw,kf. OR (functional near infrared spectroscopy or near infrared spectroscopy).tw,kf. OR (functional magnetic resonance imaging or magnetic resonance imaging).tw,kf. OR magnetoencephalograph*.tw,kf. OR transcranial magnetic stimulation.tw,kf.	Spectroscopy, Near-Infrared/ OR Magnetic Resonance Imaging/ OR Magnetoencephalography/ OR Transcranial Magnetic Stimulation/
(2) Child	child*.tw,kf. OR infan*.tw,kf. OR adolescen*.tw,kf. OR teen*.tw,kf. OR preschool.tw,kf. OR pre school.tw,kf. OR school age.tw,kf. OR youth.tw,kf. OR (school adj2 age).tw,kf. OR newborn*.tw,kf. OR kid.tw,kf. OR elementary school.tw,kf. OR middle school.tw,kf. OR high school.tw,kf. OR kindergar*.tw,kf. OR young.tw,kf.	Child/ OR Child, Preschool/ OR Adolescent/ OR infant/ OR infant, newborn/
(3) Bilingual	bilingual*.tw,kf. OR second language.tw,kf. OR ((second or dual or heritage or home or immigrant) adj2 language).tw,kf. OR dual language.tw,kf. OR heritage language.tw,kf. OR home language.tw,kf. OR L2.tw,kf. OR immigrant language.tw,kf. OR multilingual*.tw,kf.	Multilingualism/

OVID Embase

Concepts	Keywords	EMTREE Subject Headings
(1) Method	(fNIRS or NIRS).tw,kw. OR (fMRI or MRI).tw,kw. OR MEG.tw,kw. OR TMS.tw,kw. OR (functional near infrared spectroscopy or near infrared spectroscopy).tw,kw. OR (functional magnetic resonance imaging or magnetic resonance imaging).tw,kw. OR	Near-Infrared Spectroscopy/ OR Functional Near-Infrared Spectroscopy/ OR Nuclear Magnetic Resonance Imaging/ OR Magnetoencephalography/ OR Transcranial Magnetic Stimulation/

	magnetoencephalograph*.tw,kw. OR transcranial magnetic stimulation.tw,kw.	
(2) Child	child*.tw,kw. OR infan*.tw,kw. OR adolescen*.tw,kw. OR teen*.tw,kw. OR (preschool or pre school).tw,kw. OR school age.tw,kw. OR youth.tw,kw. OR (school adj2 age).tw,kw. OR newborn*.tw,kw. OR kid.tw,kw. OR elementary school.tw,kw. OR middle school.tw,kw. OR high school.tw,kw. OR kindergar*.tw,kw. OR young.tw,kw.	Child/ OR Preschool child/ OR School child/ OR Toddler/ OR Adolescence/ OR Adolescent/ OR Infant/ OR Baby/ OR Newborn/
(3) Bilingual	bilingual*.tw,kw. OR second language.tw,kw. OR ((second or dual or heritage or home or immigrant) adj2 language).tw,kw. OR dual language.tw,kw. OR heritage language.tw,kw. OR home language.tw,kw. OR L2.tw,kw. OR immigrant language.tw,kw. OR multilingual*.tw,kw.	Multilingualism/ OR Bilingualism/

OVID PsycINFO

Concepts	Keywords	APA Thesaurus of Psychological Index Terms
(1) Method	(fNIRS or NIRS).tw, OR (fMRI or MRI).tw, OR MEG.tw, OR TMS.tw, OR (functional near infrared spectroscopy or near infrared spectroscopy).tw, OR (functional magnetic resonance imaging or magnetic resonance imaging).tw, OR magnetoencephalograph*.tw, OR transcranial magnetic stimulation.tw,	Spectroscopy/ OR Magnetic resonance imaging/ OR Functional magnetic resonance imaging/ OR Magnetoencephalography/ OR Transcranial magnetic stimulation/
(2) Child	child*.tw, OR infan*.tw, OR adolescen*.tw, OR teen*.tw, OR (preschool or pre school).tw, OR school age.tw, OR youth.tw, OR (school adj2 age).tw, OR newborn*.tw, OR kid.tw, OR elementary school.tw, OR middle school.tw, OR high school.tw, OR kindergar*.tw, OR young.tw,	
(3) Bilingual	bilingual*.tw, OR second language.tw. OR ((second or dual or heritage or home or immigrant) adj2 language).tw, OR dual language.tw, OR heritage language.tw, OR home language.tw, OR L2.tw, OR immigrant language.tw, OR multilingual*.tw,	Multilingualism/ OR Bilingualism/

S2. Checklist for Quasi-Experimental Studies and Quality Appraisal Assessments

Adapted from Joanna Briggs Institute Appraisal Tools[1]

The queries from the original critical appraisal tool checklist are as follows:

Q1.   Clearly defined their inclusion criteria: criteria for inclusion in this sample should be clearly defined. The authors should provide clear inclusion/exclusion criteria developed prior to the recruitment of the participants and it should specify any necessary information critical to the study.

Q2.   Any potential confounds were considered or identified: typical confounders include baseline characteristics, prognostic factors, or concomitant exposures (i.e., maybe proficiency in our case). A confounder influences the direction of the study results. A high-quality study identified potential confounders and measured them (if possible).

Q3.   Outcomes were validly and reliably measured: methods sections of articles should include the use of objective, validated instruments and describe how the measurement was conducted (i.e., which cognitive/linguistic tasks, how they were measured, who measured them).

Q4.   Appropriate statistical analyses were used: studies should give enough detail to identify statistical technique/specific variables and how they were measured to assess appropriateness of the analysis strategy

The queries which were added by the review team are as follows:

Q5.   Measure of socioeconomic status (SES) as matched between groups: As SES is an important factor in the study of bilingualism, measuring SES and matching between groups would be ideal in the articles selected (i.e., recruitment from the same neighborhood, parental education, household income).

Q6.   Assessment for bilingual language background: studies should assess bilingual language background, specifically to find out two important areas to bilingualism: age of acquisition and proficiency.

References

Joanna Briggs Institute. Checklist for Quasi-Experimental Studies. *The Joanna Briggs Institute*. Published online 2017. Accessed March 12, 2024. https://jbi.global/critical-appraisal-tools

Quality appraisal of fMRI and fNIRS studies individual responses and summary scores

fMRI studies

	Q1	Q2	Q3	Q4	Q5	Q6	Average (SD)
Archila-Suerte 2013	Y	U	Y	Y	Y	Y	91.7 (20.4)
Ip 2017	Y	U	Y	Y	Y	Y	91.7 (20.4)
Kobayashi 2007	Y	Y	Y	Y	N	Y	83.3 (40.8)
Meng 2016	Y	Y	Y	Y	N	Y	83.3 (40.8)
Pierce 2014	Y	Y	Y	Y	Y	Y	100 (0)
Pierce 2015	Y	U	Y	Y	Y	Y	91.7 (20.4)

fNIRS studies

	Q1	Q2	Q3	Q4	Q5	Q6	Average (SD)
Abboub 2016	Y	U	Y	Y	N	N	58.3 (49.2)
Arredondo 2017	Y	U	Y	Y	Y	Y	91.7 (20.4)
Arredondo, Hu, Satterfield 2019	Y	U	Y	Y	Y	Y	91.7 (20.4)
Arredondo, Hu, Siefert 2019	Y	Y	Y	Y	Y	Y	100 (0)
Arredondo, Aslin 2022	Y	U	Y	Y	Y	N	75.0 (41.8)
Arredondo, Kovelman 2022	Y	Y	Y	Y	Y	Y	100 (0)
Baron 2023	Y	Y	Y	Y	Y	Y	100 (0)
Ding 2021	Y	Y	Y	Y	Y	Y	100 (0)
Groba 2018	Y	Y	Y	Y	N	U	75.0 (41.8)
Groba 2019	Y	Y	Y	Y	N	Y	83.3 (40.8)
Hidaka 2012	Y	U	Y	Y	U	U	75.0 (27.4)
Kou 2024	Y	Y	Y	Y	N	U	75.0 (41.8)
Li 2019	Y	Y	Y	Y	Y	U	91.7 (20.4)
Li 2020	Y	Y	Y	Y	Y	U	91.7 (20.4)
Mercure 2019	Y	Y	Y	Y	Y	N	83.3 (40.8)
Moriguchi 2020	Y	Y	Y	Y	N	U	75.0 (41.8)
Petitto 2012	Y	U	Y	Y	N	Y	75.0 (41.8)
Sun 2023	Y	Y	Y	Y	N	Y	83.3 (40.8)
Wagley 2024	Y	Y	Y	Y	Y	Y	100 (0)
Xie 2020	Y	Y	Y	Y	Y	U	91.7 (20.4)
Zhang 2023	Y	Y	Y	Y	U	Y	91.7 (20.4)

Note. N: no, NA: not applicable, U: unclear, Y: yes

Key to colour coding: Below threshold

S3. fNIRS Study Characteristics

Study	System	Wavelength (nm)	Channels	Source	Detector	S-D distance (cm)	Sampling rate (Hz)	Regions	Side	Mounting and optode placement	Dependent variables
Arredondo et al (2017)	NIRSOptix TechEN-CW6	690, 830	14	4	12	2.7	10	Prefrontal	Bilateral	Custom-built head cap, two lower sources anchored at F7 and F8	HbO, HbR
Arredondo, Hu, Satterfield et al (2019a)	NIRSOptix TechEN-CW6	690, 830	44	14	24	2.7	50	Frontal, Temporal, Parietal, Occipital	Bilateral	Cap applied using 10-10 transcranial system positioning	HbO, HbR
Arredondo, Hu, Siefert et al (2019b)	NIRSOptix TechEN-CW6	690, 830	16	4	12	2.7	10	Prefrontal	Bilateral	Custom-built head cap, two lower sources anchored at F7 and F8	HbO, HbR
Arredondo, Aslin et al (2022a)	NIRx NIRScout	760, 850	42	16	16	2 for 42- and 44 cm caps, 2.25 for 46- and 48 cm caps	3.91	Frontal, temporal, parietal	Bilateral	EasyCap (a 42-cm, 44-cm, 46-cm, or 48-cm EEG head cap)	HbO, HbR
Arredondo, Kovelman et al (2022b)	NIRSOptix TechEN-CW6	690, 830	44	14	24	2.7	50	Frontal, Temporal, Parietal, Occipital	Bilateral	Cap applied using 10-10 transcranial system positioning	HbO, HbR
Baron et al (2023)	NIRSOptix TechEN-CW6	690, 830	16	5	8	2.7	50	Frontal, Temporal	Left	Custom built cap 10–10 transcranial system positioning	HbO, HbR
Ding et al (2021)	Hitachi ETG-4000	695, 830	44	16	14	n.r.	10	Frontal, temporal	Mostly left	Probe caps two arrays 3x5	AUC
Groba et al (2018)	NIRx NIRScout	760, 850	26	9	14	2.5	6.25	Prefrontal, frontal, temporal, parietal	Bilateral	Modified Easy Cap	HbO, HbR
Groba et al (2019)	NIRx NIRScout	760, 850	26	9	14	2.5	6.25	Prefrontal, frontal,	Bilateral	Modified Easy Cap	HbO, HbR

Study	Device	Wavelengths						Region	Laterality	Details	Chromophore
Hidaka et al (2012)	Hitachi ETG-100	780, 830	24	8	10	3	10	Temporal	Bilateral	2 3x3 matrices probes (vertical centers aligned to vertex and mastoid tip; bottom 3 probes aligned with nasion and inion)	HbO, HbR
Kou et al (2024)	NIRx NIRScout	760, 850	39	16	16	n.r.	3.91	Frontal, Parietal, Temporal	Bilateral	fNIRS cap. 10–20 system positioning.	HbO, HbR
Li et al (2019)	NIRx NIRSport	630, 850	20	8	8	n.r.	7.81	Prefrontal	Bilateral	Detector optode 1 and 5 anchored at F7 and F8 using international 10–10 transcranial system positioning. The source optode 2 and 6 were anchored at F5 and F6.	HbO, HbR
Li et al (2020)	NIRx NIRSport	630, 850	20	8	8	2.5	7.81	Prefrontal	Bilateral	Per the 10–10 transcranial positioning system, four reference detectors (1, 3, 7, and 5) were positioned at F7, F3, F4, and F8.	HbO, HbR
Mercure et al (2019)	Gowerlabs NTS optical topography system	770, 850	46	16	16	2	10	Frontal, temporal, temporo-parietal	Bilateral	Custom-built fNIRS headgear	HbO, HbR
Moriguchi & Lertladaluck (2020)	Spectratech OEG-16	770, 840	16	6	6	3	6.10	Prefrontal	Bilateral	Lower row of the fNIRS probe was spread between F7 and F8, whereas the center point of the probe was placed at Fpz	HbO, HbR
Petitto et al (2012)	Hitachi ETG-4000	690, 830	24	5	4	n.r.	10	Frontal, Temporal	Bilateral	Secured using a terrycloth headband. 2 3x3 arrays, positioned using 10-20 system, lowest optode rows placed over T3/T4	HbO, HbR
Sun et al (2023)	NIRSOptix TechEN-CW6	690, 830	46	12	24	2.7	50	Frontal, Temporal, Parietal	Bilateral	F7, F8, T3, and T4 were anchored to respective	HbO

										sources/detectors on fNIRS cap	
Wagley et al (2024)	NIRSOptix TechEN-CW6	690, 830	16	5	8	2.7	50	Frontal, Temporal	Bilateral	Two custom-built head caps, 54 cm and 56 cm. Probe placements were based on MRI and photogrammetry-based stereoscopic optode registration.	HbO, HbR
Xie et al (2020)	Oxymon Mk III	760, 850	17	7	7	2.5	50	Prefrontal, Frontal	Mostly left	Upper central probe, which was anchored at Fz according to 10–20 system.	HbO
Zhang et al (2023)	NIRSOptix TechEN-CW6	690, 830	36	12	18	2.7	50	Frontal, Temporal	Bilateral	F7, F8, T3, and T4 were anchored to a specific source or detector using silicone band fNIRs caps	HbO, HbR

Note. AUC - area under the curve; HbO - oxygenated hemoglobin; HbR - deoxygenated hemoglobin; n.r. – not reported

S4. Probability Maps Meta-Analysis Approach

Methods

To further characterize the functional significance of ALE clusters identified in our meta-analysis, we adopted a probabilistic functional atlas approach (Hauptman et al., 2023). This method estimates the probability that specific brain coordinates belong to functionally predefined networks.

This exploratory analysis was applied to two analysis sets: (1) within-group bilingual contrasts (Set 1), which yielded 7 significant clusters, and (2) the bidirectional bilingual-monolingual contrast (Set 2B), which yielded 1 significant cluster in the left inferior frontal gyrus (IFG). The unidirectional bilingual > monolingual contrast (Set 2A) yielded no significant clusters and was therefore not included in this analysis.

Probabilistic Functional Atlases

We evaluated cluster coordinates against three probabilistic network atlases: Language (LangA; Lipkin et al., 2022), Theory of Mind (ToM) and Multiple Demand (MD; Lipkin et al., 2023). Full methodological details are available in the original publications and methodological supplements. These atlases were constructed by selecting the top 10% of most localizer-responsive voxels for each participant (to account for inter-individual variability in activation magnitude), then overlaying these individual binarized activation maps in common MNI space. The resulting value at each voxel represents the proportion of participants for whom that voxel belonged to the network, serving as an estimate of network probability.

Network Probability Extraction

For each cluster, we conducted two complementary analyses:

- **Peakwise Analysis (Center Peak):** We extracted network probabilities for the center peak coordinate (the voxel with the highest ALE value) of each significant cluster. Where there were two peaks for a single cluster, both were characterized. This provides an estimate of the functional network affiliation of the most convergent activation location within each cluster.

- **Voxelwise Analysis (All Voxels in a Cluster):** We extracted network probabilities for all voxels contained within each significant cluster and calculated the median network probability across all voxels for each network. This approach provides a more comprehensive characterization of the functional network profile by considering the full spatial extent of each cluster rather than relying solely on a single peak coordinate.

Statistical Testing

To assess statistical reliability, we implemented a permutation-based procedure (Hauptman et al., 2023). For each cluster, we calculated observed median differences in network probabilities between each network pair (Language vs. ToM, Language vs. MD, ToM vs. MD). Hemisphere-specific null distributions were generated by randomly sampling voxel coordinates from 3D probability maps (excluding zero-probability voxels in all three networks) matching observed cluster sizes, repeated 10,000 times. We used one-tailed tests for directional hypotheses based on prior literature on functional network properties (Fedorenko & Blank, 2020; Hauptman et al., 2023).

Results

In general, activation patterns showed the strongest correspondence with the Language network, particularly in left hemisphere regions, with more modest associations with the MD and ToM networks. These findings suggest that observed activation patterns in bilingual

children reflect primarily language-specific neural processing, with additional contributions from

domain-general executive control in specific anatomical regions (cerebellar, parietal), but

minimal involvement of social-cognitive (ToM) mechanisms.

Set 1 (Bilingual Only)

Overall network probabilities for Set 1 for peakwise and voxelwise analysis are provided

in Table S4.1.

Table S4.1.

Network probabilities for the seven significant ALE clusters of Set 1 (Bilingual only) contrasts.

Cluster	Size (voxels)	Anatomical area	MNI Coordinates (x, y, z)			Language (peak / voxel median)	ToM (peak / voxel median)	MD (peak / voxel median)
1	236	RH STG	62	-10	2	**.078/.092**	.020/.025	.023/.016
2	198	LH STG	-58	-22	8	**.081/.134**	.015/.023	.009/.010
			-60	-14	4	**.138/.134**	.016/.023	.015/.010
3	50	RH Precentral gyrus	54	-2	48	**. 329/.274**	.049/.045	.035/.043
4	42	LH Culmen	-28	-62	-26	.134/.095	.066/.068	**.217/.310**
5	34	LH STG	-42	-40	8	**.163/.182**	.025/.025	.004/.007
6	27	LH SMG	-42	-44	42	.051/.051	.010/.010	**.631/.592**
7	27	LH MFG / Cingulate gyrus	-12	30	44	**.048/.032**	.040/.040	.023/.281
			-6	26	42	.031/.032	.045/.040	**.433/.281**

Note. The network with the highest probability for each cluster is indicated in bold. ToM: Theory of mind; MD: Multiple demand; RH = Right Hemisphere; LH = Left Hemisphere; STG = Superior Temporal Gyrus; SMG = Supramarginal gyrus; MFG = Middle Frontal Gyrus.

Peakwise Analysis. Examining the center peak of each cluster revealed a median

network probability of .082 for the Language network, substantially higher than ToM (.025) and

MD (.023). However, this general pattern masked considerable hemispheric and anatomical

heterogeneity. In the left hemisphere, permutation testing demonstrated that center peaks were

located significantly more centrally in the Language network than in the ToM network (observed

median difference = 0.042, p = .008), supporting the hypothesis that left-lateralized regions in bilingual children preferentially engage language-specific processing mechanisms. The Language network exhibited highest probabilities in temporal regions (LH STG clusters, peaks ranging from .081 to .163) and in the RH precentral gyrus (.329), suggesting robust language network engagement across multiple functional-anatomical systems.

In contrast, the MD network showed the highest probabilities in cerebellar (culmen: .217), parietal (supramarginal gyrus: .631), and medial frontal regions (MFG/cingulate: .433 at one peak), consistent with supporting domain-general cognitive control and working memory processes. In the right hemisphere, center peaks showed significantly higher MD than ToM probabilities (observed median difference = 0.143, p < .0001), suggesting that right-lateralized activation in bilingual children may reflect executive processing demands.

Voxelwise Analysis. Voxels patterns of network dominance were consistent with peakwise analyses. The Language network continued to predominate in temporal and precentral clusters (Clusters 1, 2, 3, 5) and the MD network showed even stronger dominance in cerebellar, parietal, and medial frontal regions (Clusters 4, 6, 7). Permutation testing on the median voxelwise probabilities confirmed significant Language > ToM differences in the left hemisphere (p = .002) and significant MD > ToM differences in both hemispheres (LH: p < .001; RH: p < .001).

Set 2 (Bidirectional Bilingual ≠ Monolingual)

Overall network probabilities for Set 2B for peakwise and voxelwise analysis are provided in Table S4.2.

Table S4.2.

Network probabilities for the single significant ALE cluster of Set 1 (Bilingual only) contrasts.

Cluster	Size (voxels)	Anatomical Area	MNI Coordinates (x, y, z)			Language Network Probability (center peak / voxel median)	ToM Network Probability (center peak / voxel median)	MD Network Probability (center peak / voxel median)
1	57	LH IFG	-44	24	0	**0.383 / 0.274**	0.191 / 0.126	0.060 / 0.047

Note. The network with the highest probability for each cluster is indicated in bold. ToM: Theory of mind; MD: Multiple demand; LH = Left Hemisphere; IFG = Inferior Frontal Gyrus.

Peakwise Analysis. The center peak of the left IFG cluster exhibited a median network probability of .383 for the Language network, substantially higher than ToM (.191) and MD (.061). Permutation testing confirmed that this Language network dominance was statistically robust, with the Language network showing significantly higher probabilities than both MD (observed median difference = 0.322, p = .033) and ToM (observed median difference = 0.191, p = .034). No significant difference emerged between ToM and MD networks (p = .907).

Voxelwise Analysis. Convergent results were seen in the voxel patterns, with even stronger statistical evidence. The median voxelwise network probability was .274 for Language, .126 for ToM, and .048 for MD. Permutation testing demonstrated highly significant Language > MD (observed median difference = 0.217, p < .0001) and Language > ToM differences (observed median difference = 0.164, p < .0001), with no significant MD-ToM difference (p = 1.0).

Synthesis Across Analysis Sets

Taken together, these findings reveal distinct but complementary functional network profiles across within-group bilingual activations and between-group bilingual-monolingual differences. Within bilingual children (Set 1), activation patterns engage both the Language

network (left temporal regions) and the MD network (cerebellar, parietal, and frontal regions) suggesting that bilingual children recruit both language-specific and domain-general executive mechanisms during experimental tasks.

In contrast, the single significant cluster differentiating bilingual from monolingual children (Set 2B) showed a clear Language network profile (with minimal MD or ToM involvement), suggesting that group differences are driven primarily by language-specific processing mechanisms. This pattern aligns with perspectives emphasizing the importance of language network adaptations in bilingual development, while also highlighting that within-group activation patterns (which may reflect task performance demands) can engage broader networks beyond those showing group differences.

The ToM network exhibited consistently low probabilities across both analysis sets, despite inclusion of social reasoning tasks. This suggests minimal overlap between our meta-analytic peaks and the ToM network as defined by false belief localizers, potentially reflecting differences in task demands, developmental stage, or spatial localization in pediatric populations.

References

Fedorenko, E., & Blank, I. A. (2020). Broca's Area Is Not a Natural Kind. *Trends in Cognitive Sciences*, *24*(4). https://doi.org/10.1016/j.tics.2020.01.001

Hauptman, M., Blank, I., & Fedorenko, E. (2023). Non-literal language processing is jointly supported by the language and theory of mind networks: Evidence from a novel meta-analytic fMRI approach. *Cortex*, *162*, 96–114. https://doi.org/10.1016/j.cortex.2023.01.013

Lipkin, B., Blank, I., & Fedorenko, E. (2023). *Lipkin et al. (2022) Online Supplement: Probabilistic atlases for the multiple demand (MD) and theory of mind (ToM) networks based on large-scale precision localizers.* 8503288 Bytes. https://doi.org/10.6084/M9.FIGSHARE.22306348.V2

Lipkin, B., & Tuckute, G. (2022). *LanA (Language Atlas): A probabilistic atlas for the language network based on fMRI data from 800+ individuals.* https://doi.org/10.17605/OSF.IO/KZWBH