Table S1. Full sample standardization constants

Task	Mean	Sample SD
RMET	23.766667	3.250684
Facial emotion recognition	48.366667	3.412615
Faux Pas	39.833333	3.184345

Table S2. Observations flagged by residual screening

Dataset ID	Model	Studentized residual	Cook's D	Leverage	Group DFBETA
5	RMET	−4.648	0.1327	0.0680	−0.345
1	Facial emotion	−3.016	0.0648	0.0710	−0.305

Note. Both participants were bilingual. Cook's D, leverage, and group DFBETAs were reported descriptively and were not used as exclusion criteria.

Table S3. Participant characteristics by recruitment site

Participant characteristics	Kars (n = 60)	Mardin (n = 60)	*p*
Age, mean (SD) [a]	15.00 (1.98)	15.20 (2.03)	0.586
Sex, % (n) [b]			0.465
Male	48.3 (29)	55.0 (33)	
Female	51.7 (31)	45.0 (27)	
School stage, % (n) [b]			0.190
Middle school	45.0 (27)	33.3 (20)	
High school	55.0 (33)	66.7 (40)	
Monthly household income, % (n) [b]			0.180
Low	30.0 (18)	23.3 (14)	
Middle	40.0 (24)	56.7 (34)	
High	30.0 (18)	20.0 (12)	
Maternal education, % (n) [b]			0.798
Primary school	16.7 (10)	18.3 (11)	
Middle school	25.0 (15)	28.3 (17)	
High school	25.0 (15)	28.3 (17)	
University	33.3 (20)	25.0 (15)	
IQ score, mean (SD) [a]	101.38 (14.17)	102.75 (13.01)	0.583

Note. SD: standard deviation; IQ: intelligence quotient. [a] Independent-samples t test. [b] Pearson's chi-square test.

Table S4. Residual distribution and variance diagnostics

Sample	Outcome	n	Shapiro–Wilk *p*	Levene *p*
Full sample	RMET	120	0.022	0.657
Full sample	Facial emotion	120	0.565	0.738
Full sample	Faux Pas	120	0.052	0.947
Two observations excluded	RMET	118	0.927	0.989
Two observations excluded	Facial emotion	118	0.789	0.849
Two observations excluded	Faux Pas	118	0.050	0.996

Table S5. Tests of common age and IQ slopes

Model / outcome	Group x age *p*	Group x IQ *p*	Joint HC3 test *p*
MANCOVA	0.144	0.478	—
RMET	0.340	0.130	0.259
Facial emotion recognition	0.139	0.751	0.282
Faux Pas	0.347	0.869	0.579

Note. Joint tests evaluated whether the group x age and group x IQ coefficients were both zero.

Table S6. Multivariate language-group effects

Model	n	Pillai's trace	F	df	*p*
Full sample	120	0.395	23.46	3,108	< 0.001
Two observations excluded	118	0.388	22.37	3,106	< 0.001
Quadratic age added	120	0.400	23.82	3,107	< 0.001
Group-specific age and IQ slopes	120	0.400	23.51	3,106	< 0.001
Ages 12–16 years (WISC-R subgroup)	83	0.512	24.83	3,71	< 0.001
Recruitment site-adjusted model	120	0.400	23.73	3,107	< 0.001

Table S7. Adjusted unstandardized group differences

Model	Outcome	B [95% CI]	Holm *p*
Full sample	RMET	−3.04 [−4.18, −1.91]	< 0.001
	Facial emotion	−3.54 [−4.68, −2.40]	< 0.001
	Faux Pas	0.01 [−1.23, 1.25]	0.986
Two observations excluded	RMET	−2.91 [−4.00, −1.81]	< 0.001
	Facial emotion	−3.42 [−4.53, −2.32]	< 0.001
	Faux Pas	0.12 [−1.12, 1.36]	0.851
Quadratic age added	RMET	−3.09 [−4.23, −1.95]	< 0.001
	Facial emotion	−3.56 [−4.71, −2.42]	< 0.001
	Faux Pas	0.04 [−1.21, 1.29]	0.952
Group-specific age and IQ	RMET	−3.00 [−4.14, −1.86]	< 0.001
slopes	Facial emotion	−3.53 [−4.66, −2.40]	< 0.001
	Faux Pas	0.02 [−1.25, 1.29]	0.975
Ages 12–16 years	RMET	−3.63 [−5.01, −2.26]	< 0.001
(WISC-R subgroup)	Facial emotion	−4.29 [−5.62, −2.97]	< 0.001
	Faux Pas	0.06 [−1.33, 1.45]	0.932
Recruitment site-adjusted model	RMET	−3.01 [−4.13, −1.89]	< 0.001
	Facial emotion	−3.52 [−4.67, −2.38]	< 0.001
	Faux Pas	0.01 [−1.26, 1.27]	0.994

Note. B represents the bilingual-minus-monolingual difference. Confidence intervals were not adjusted for multiple comparisons.

Table S8. Group contrasts and interaction tests

Model	n	RMET Estimate [95% CI]	Facial emotion Estimate [95% CI]	Faux Pas Estimate [95% CI]	χ^2 / p
Full sample	120	−0.94 [−1.25, −0.62]	−1.04 [−1.34, −0.73]	0.00 [−0.35, 0.36]	22.43 < 0.001
Two observations excluded	118	−0.89 [−1.20, −0.59]	−1.00 [−1.30, −0.71]	0.04 [−0.31, 0.39]	21.47 < 0.001
Quadratic age added	120	−0.95 [−1.26, −0.64]	−1.04 [−1.35, −0.74]	0.01 [−0.34, 0.36]	23.68 < 0.001
Group-specific age and IQ slopes	120	−0.92 [−1.24, −0.61]	−1.03 [−1.33, −0.74]	0.01 [−0.35, 0.36]	22.46 < 0.001
Ages 12–16 years (WISC-R subgroup)	83	−1.12 [−1.48, −0.75]	−1.26 [−1.60, −0.92]	0.02 [−0.36, 0.39]	29.62 < 0.001
Recruitment site-adjusted model	120	−0.93 [−1.23, −0.62]	−1.03 [−1.33, −0.73]	0.00 [−0.35, 0.36]	21.90 < 0.001
Independent correlation	120	−0.94 [−1.25, −0.62]	−1.04 [−1.34, −0.73]	0.00 [−0.35, 0.36]	22.43 < 0.001

Note. χ^2 and *p* values refer to the group x task interaction (2 df). Confidence intervals were not adjusted for multiple comparisons.

Table S9. Differences between task-specific group contrasts

Model	RMET minus Faux Pas Estimate [95% CI]	Facial emotion minus Faux Pas Estimate [95% CI]	RMET minus facial emotion Estimate [95% CI]
Full sample	−0.94 [−1.40, −0.48]	−1.04 [−1.51, −0.57]	0.10 [−0.33, 0.53]
Two observations excluded	−0.93 [−1.38, −0.48]	−1.04 [−1.52, −0.56]	0.11 [−0.30, 0.51]
Quadratic age added	−0.96 [−1.42, −0.50]	−1.06 [−1.52, −0.59]	0.09 [−0.34, 0.53]
Group-specific age and IQ slopes	−0.93 [−1.39, −0.47]	−1.04 [−1.50, −0.58]	0.11 [−0.32, 0.54]
Ages 12–16 years (WISC-R subgroup)	−1.14 [−1.60, −0.67]	−1.28 [−1.81, −0.75]	0.14 [−0.35, 0.63]
Recruitment site-adjusted model	−0.93 [−1.39, −0.47]	−1.03 [−1.50, −0.56]	0.11 [−0.32, 0.54]

Note. All RMET–Faux Pas and facial emotion recognition–Faux Pas contrasts had Holm-adjusted $p < .001$. RMET–facial emotion recognition contrasts were nonsignificant. Confidence intervals were not adjusted for multiple comparisons.