Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Effects of Foreign Language Learning on Cognitive Function in Healthy Older Adults

Wei Zhang
Hainan University
Jingjing Chen
Hainan University
Yanting Wang
Qiongtai Normal University
Fengru Zeng
Hainan University
Fang Ouyang
Hainan University
Yi Liao
liaoy19@tsinghua.org.cn

Qiongtai Normal University

Article

Keywords: foreign language learning, cognitive function, healthy older adults, meta-analysis, randomized controlled trial

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11184942/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

The Effects of Foreign Language Learning on Cognitive Function in Healthy Older Adults

Wei Zhang[1,2], **Jingjing Chen**[1*], **Yanting Wang**[3*], **Fengru Zeng**[1], **Fang Ouyang**[1], **Yi Liao**[3,4*]

[1] School of International studies, Hainan University, Haikou, China

[2] Institute of Language Cognition, Carleton University, Ottawa, Canada

[3] School of Arts, Qiongtai Normal University, Haikou, China

[4] Institute for Interdisciplinary Studies, McMaster University, Hamilton, Canada

*Corresponding author

Corresponding author1: Jingjing Chen

E-mail: 87270273@qq.com

Postal Address: Hainan University, 58 Renmin Rd., Haikou, Hainan, China 570228

Affiliation: School of International studies, Hainan University, Haikou, China

Corresponding author2: Yanting Wang

E-mail: 2552767741@qq.com

Postal Address: Guilinyang University Town, Haikou China 571127

Affiliation: School of Arts, Qiongtai Normal University, Haikou, China

Corresponding author3: Yi Liao

E-mail: liaoy19@tsinghua.org.cn

Postal Address: Guilinyang University Town, Haikou China 571127

Affiliation: School of Arts, Qiongtai Normal University, Haikou, China

Abstract

This meta-analysis synthesised randomised controlled trial (RCT) evidence on whether structured foreign-language learning improves cognitive performance in cognitively healthy older adults. PubMed, Cochrane Library, Embase and Web of Science were searched from inception to 14 March

2026. Pooled effect sizes were calculated as Hedges' g under random-effects models, with exploratory mixed-effects subgroup analyses used to examine intervention- and participant-level moderators. Eleven eligible RCTs comprising 840 older adults were included. Foreign-language learning was associated with a small-to-moderate pooled estimate of overall cognitive benefit (Hedges'g = 0.309, 95% CI 0.153–0.465, $p < 0.001$; $I^2 = 55\%$), alongside a significant pooled effect on global cognition. Because the primary estimate pooled heterogeneous cognitive outcomes, it should be interpreted as a broad summary rather than a unitary construct. Subgroup differences were exploratory and based on small subgroups; they indicated larger estimates for cross-family language pairs, moderate-to-high weekly training and online self-directed learning, whereas age, sex and education showed no significant moderating effects. Evidence certainty was rated low-to-moderate, driven mainly by performance bias and between-study heterogeneity. Findings suggest that foreign-language learning may represent a promising, though not yet established, non-pharmacological cognitive enrichment option for older adults; benefits, if present, appeared to be shaped more by intervention implementation than by demographic traits. Further large, pre-registered RCTs with active controls and long-term follow-up are required.

Keywords: foreign language learning; cognitive function; healthy older adults; meta-analysis; randomized controlled trial

1. Introduction

Global population aging is accelerating, and age-related cognitive decline creates substantial barriers to older adults' independence while placing heavy burdens on public health systems worldwide. Dementia prevalence continues to rise across the globe; the Alzheimer's Association (2024) estimates that roughly 55 million people currently live with dementia, and this figure is projected to grow markedly in coming decades. Even among cognitively intact seniors, normal aging brings measurable declines in attentional control, working memory, executive function, and processing speed, which can undermine daily functioning and social participation (Buckner, 2004; Diamond, 2013). Mild cognitive impairment, a transitional state between healthy aging and dementia, affects 15-20 % of

adults aged 65 years and above and elevates subsequent dementia risk (Petersen, 2011; Gonzales et al., 2022). Against this backdrop, identifying safe, accessible non-pharmacological approaches to preserve cognitive health in later life has become a key priority for neuroscience, psychology, and geriatric public health research.

Cognitive reserve theory provides a core theoretical framework for understanding cognitive resilience in old age. Repeated engagement in intellectually stimulating activities can strengthen the brain's capacity to withstand age-related neural changes via flexible, efficient cognitive processing (Stern, 2009; Valenzuela & Sachdev, 2006). Importantly, cognitive enrichment rarely stems from a single behaviour. Intellectual effort, physical activity, and social interaction interact to shape late life cognition through sustained mental engagement and skill acquisition (Hertzog et al., 2008). Previous meta analyses focusing on non-pharmacological cognitive interventions have demonstrated that intervention outcomes are heavily contingent upon modifiable implementation characteristics such as training intensity, delivery format, and supervision (Lampit et al., 2014; Ngandu et al., 2015). As a complex mental activity requiring simultaneous deployment of attention, memory updating, information processing, and rule based reasoning, structured foreign-language learning has emerged as a promising candidate cognitive enrichment activity for older adults.

Much of the initial interest in language driven cognitive benefits originates from observational bilingual cognition research. Lifelong bilingual experience demands constant monitoring, selection, and inhibition across competing language systems, which may engage executive control and frontal lobe brain networks over the lifespan. Early observational work linked lifelong bilingualism to delayed dementia onset and better cognitive maintenance in older populations (Bialystok et al., 2012; Paulavicius et al., 2020). Nevertheless, the so-called bilingual advantage remains contested. Meta-analytic syntheses have documented heterogeneous effect sizes that are sensitive to task demands, participant backgrounds, and language use histories, with no consistently replicable universal cognitive benefit (Adesope et al., 2010; Lehtonen et al., 2018; Grundy, 2020). Critically, correlational findings from lifelong bilingual samples cannot be directly transferred to older adults undertaking new structured foreign language training.

Lifelong bilingualism and late adulthood foreign language learning represent fundamentally distinct forms of language experience. Lifelong bilingualism typically originates in childhood or early adulthood through extensive, naturalistic real world language exposure. By contrast, foreign language learning initiated after age 60 is generally time limited, formally structured, and strongly shaped by instructional design, learning resources, and participant engagement patterns. Consequently, cognitive patterns observed among long term bilingual individuals cannot be assumed to occur in seniors who commence language study late in life; randomised controlled trials are required to test whether structured foreign language learning itself produces measurable cognitive gains independent of pre-existing individual differences. Two prominent neurocognitive theories support the theoretical plausibility of late-life learning related plasticity. The Scaffolding Theory of Aging and Cognition (STAC) posits that the aging brain can sustain function by recruiting additional neural resources and building compensatory brain networks (Park & Reuter-Lorenz, 2009). Complementing this, the Dynamic Restructuring Model (DRM) argues that novel language learning can trigger structural and functional neural adaptation well into older age, demonstrating preserved neuroplastic potential (Pliatsikas, 2020). Drawing on these models, several scholars have proposed that late-life foreign language learning constitutes valuable cognitive enrichment by taxing attention, memory, executive control, and learning related mechanisms concurrently (Antoniou et al., 2013). Still, whether these cognitively demanding learning experiences translate into detectable behavioural cognitive improvements remains an open empirical question requiring rigorous intervention-based evaluation.

Existing RCT evidence regarding the cognitive impacts of late-life foreign language learning is inconsistent. Some trials have reported positive cognitive outcomes following structured language interventions. Wong et al. (2019) observed broad globa cognition improvements after six month English training, and Grossmann et al. (2023) found enhanced response inhibition among older participants with lower baseline cognition following intensive Spanish instruction. Digital platform supported language learning studies also suggest technology-mediated practice may foster executive function related cognitive engagement (Parmaxi & Zaphiris, 2017; Chan et al., 2022). Conversely, multiple conducted RCTs have yielded null or limited between-group cognitive differences. Berggren

et al. (2020) detected no cognitive advantage for language learning relative to an active relaxation control; Brouwer et al. (2025) reported minimal cognitive change from English training compared with a music-based control condition. Mixed or negative findings have additionally been documented, including stronger improvements within control groups on selected outcomes and decoupling between neural functional connectivity shifts and behavioural cognitive test results (Meltzer et al., 2023; Bubbico et al., 2019). Such divergent results imply that cognitive outcomes are not determined solely by study participation, but vary according to intervention intensity, design, and implementation.

Following established meta-analytic frameworks for older adult cognitive interventions (Lampit et al., 2014; Ngandu et al., 2015), we pre-specified twelve candidate moderators for subgroup analysis. These moderators were chosen based on three criteria: theoretical grounding derived from cognitive reserve and neuroplasticity models, consistent reporting across available RCTs, and emphasis within prior systematic reviews of cognitive training and language-related plasticity. Moderators fall into three categories: intervention level variables (L1-L2 linguistic distance, total intervention duration, training frequency, teacher-led versus self-directed format, delivery modality, self-study frequency, inclusion of fMRI measurement); participant level demographic variables (age classification, sex distribution, educational attainment); and one outcome measurement-related variable (global cognition assessment type). Constructs inconsistently reported across primary studies, such as personality traits and socioeconomic status beyond educational attainment, were excluded from formal testing. Several moderators carry competing theoretical predictions. Greater linguistic distance between native and target languages may increase cognitive processing load and boost neuroplastic stimulation, yet excessive dissimilarity could raise frustration and reduce adherence (Carthery-Goulart et al., 2023; Ware et al., 2021). Weekly training frequency may follow a dose response relationship, though very high training loads risk cognitive fatigue. Self-directed learning imposes high metacognitive demands that may amplify cognitive benefits, whereas teacher-led formats can lower learner burden and improve retention. Online delivery enables adaptive difficulty and real time feedback, while offline classroom learning provides rich social contact. Demographic moderators may shape intervention response, but prior meta-analyses yield inconsistent evidence for

their effects. Testing this set of moderators helps disentangle whether cross-study heterogeneity arises primarily from adjustable intervention features, stable participant demographics, or methodological measurement differences.

Despite growing research interest, three critical research gaps persist. First, existing reviews often combine evidence from observational lifelong bilingualism research and intervention based language learning studies, despite their fundamentally distinct research questions. Second, no published meta-analysis has exclusively synthesised RCT evidence focusing on cognitively healthy older adults undertaking structured foreign language learning, leaving the magnitude, reliability, and domain specific pattern of cognitive benefits poorly characterised. Third, systematic characterisation of intervention- and participant-level moderators remains absent. We adopt a conceptual framework in which foreign-language learning functions as a cognitively enriching activity rather than guaranteeing cognitive enhancement. Repeated exposure to complex cognitive challenges during language learning creates opportunities for cognitive adaptation and neuroplastic change; the magnitude of resulting behavioural cognitive effects is conditioned heavily by intervention-specific characteristics. Building upon this framework, the present systematic review and meta-analysis addresses three prespecified research questions:

RQ1: Does foreign language learning improve overall cognitive function in healthy older adults?

RQ2: Are cognitive benefits observed consistently across different cognitive domains, including global cognition, executive function, working memory, and attention?

RQ3: Which participant-related and intervention-related characteristics moderate the cognitive effects of foreign language learning interventions?

By addressing these questions, this review aims to clarify whether foreign language learning can serve as a feasible cognitive intervention for healthy aging and to identify the conditions under which such interventions may be most beneficial.

2. Materials and Methods

2.1 Study Design and Reporting Standards

This systematic review and meta-analysis evaluated the effects of foreign-language learning interventions on cognitive function in healthy older adults and explored potential moderators. A protocol specifying eligibility criteria, inclusion and exclusion criteria, data-extraction procedures and the statistical analysis strategy was developed before screening. Because the review was based on secondary analysis of published data and did not involve individual private information, no additional ethical approval was required. The review was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement. Of note, the protocol was not prospectively registered (e.g., in PROSPERO); this is acknowledged as a limitation in the Discussion.

2.2 Search Strategy

PubMed, Cochrane Library, Embase and Web of Science were systematically searched from database inception to 14 March 2026, with the last search executed on 14 March 2026.

The search strategy was constructed around the review question and combined Medical Subject Headings (MeSH) with free-text terms. Three concept blocks were used: (1) language learning terms, including language learning, language training, language course, language acquisition, foreign language and language teaching; (2) cognitive-function terms, including cognitive control, cognitive functions, cognitive flexibility, cognitive performance, cognitive decline, executive control, executive functions, executive flexibility, executive performance and executive decline; and (3) population terms, including older adult, senior, third-age, third age, aged, elderly and pension. Terms within each block were connected with OR, and the blocks were connected with AND. Reference lists of included studies and relevant reviews were searched manually. Complete database-specific search strings, formal search dates and the total number of records retrieved from each database are fully documented in Supplementary Table S1, in compliance with PRISMA 2020 transparency standards. The supplementary table is structured with four columns: Database, Search date, Full search strategy, and Number retrieved.

Two reviewers independently conducted the search and initial screening. Disagreements were resolved through discussion or adjudication by a third reviewer.

2.3 Inclusion and Exclusion Criteria

Inclusion criteria were defined according to the PICOS framework (Population, Intervention, Comparator, Outcome and Study design): (1) RCT design; (2) cognitively healthy adults aged 60 years or older, without clinically diagnosed neurological disease or cognitive impairment; (3) structured foreign language learning courses as the intervention, including listening, speaking, reading, writing or vocabulary learning; (4) no-intervention, wait-list, usual-activity or active controls such as music, meditation or cognitive games; and (5) at least one objective measure of cognitive function.

All records were cross-checked for sample overlap during full-text screening. Only independent study cohorts were included to avoid double-counting participants and effect size bias. No overlapping samples were identified across the final included studies.

Exclusion criteria were: (1) non-randomized controlled studies; (2) cross-sectional studies, cohort studies or case reports; (3) participants younger than 60 years; (4) inclusion of participants with mild cognitive impairment or dementia where data for healthy older adults could not be extracted separately; (5) interventions involving native-language training, speech therapy or rehabilitation for language disorders; (6) conference abstracts, commentaries or theoretical articles; and (7) insufficient data to calculate effect sizes, even after contacting authors.

2.4 Data Extraction

Two reviewers independently extracted data using a standardized form. Extracted information included first author, publication year, sample size, participant demographics (age, sex and education), native language (L1), target language (L2), baseline language proficiency, intervention protocol (total duration, training frequency, delivery format, teacher-led versus self-directed format and self-study requirements), control condition, cognitive outcomes and instruments, and pre- and post-intervention means and standard deviations or other statistics convertible to effect sizes. For studies with multiple time points, only immediate post-intervention assessments were extracted to maximize comparability. To standardize the heterogeneous statistical reporting across included studies and ensure the consistency and comparability of effect-size calculations, a pre-specified

hierarchical data conversion rule was implemented during data extraction, with full details of all decision logic documented in Supplementary Table S2(Data Extraction and Conversion Rules) organized by data situation and corresponding conversion approach. Specifically, when pre- and post-intervention means and standard deviations (SDs) were available for both the intervention and control groups, these values were used directly to compute change-score-based effect sizes; for studies adopting a posttest-only design that only reported post-intervention means and SDs, between-group effect sizes were estimated directly from post-intervention statistics; for outcomes reported as median with interquartile range (IQR) or median with range, the conversion method proposed by Hozo et al. (2005) was applied to derive estimated means and corresponding SDs; when pre-post correlation coefficients were not reported in original studies, the correlation was set to r = 0.5 following the guidance of the Cochrane Handbook (Furlan et al., 2009) to ensure consistent variance estimation for change-score effect sizes; for studies reporting multiple cognitive measures within a single cognitive domain, the prespecified primary outcome was prioritized for effect-size calculation, and where no primary outcome was explicitly defined, the most widely used and representative measurement across the included literature was selected to mitigate outcome reporting bias and unit-of-analysis errors. All effect sizes across different input data formats were uniformly quantified as Hedges' g with small-sample bias correction, with change- score estimates serving as the primary calculation approach to account for baseline differences and improve statistical power, and the robustness of these statistical assumptions was further evaluated through sensitivity analyses to assess their potential impact on the pooled effect estimates.

2.5 Effect-size Calculation

The standardized mean difference (SMD) was used as the primary effect metric, with Hedges' g applied to correct for small-sample bias.

Effect sizes were primarily calculated from the difference in pre-post change between intervention and control groups, as change-score estimates account for baseline differences and generally provide greater statistical power. For studies with only post-intervention data, between-group SMDs were computed directly from posttest means and pooled SDs, then corrected to

Hedges' g using the same small-sample correction factor. All conversions from median/IQR to mean/SD were completed before effect-size computation, ensuring all effect estimates were on a uniform, comparable metric.

The small-sample correction factor J was applied to all SMD estimates to derive Hedges' *g*, regardless of input data format. The core calculation formula is retained below:

$$g = J \times \frac{(M_{post} - M_{pre})_{intervention} - (M_{post} - M_{pre})_{control}}{SD_{pooled}}$$

Equation 1. Calculation of Hedges' *g* from standardized pre-post group differences.

In this equation, J denotes the Hedges small-sample correction factor. When a study did not report the pre-post correlation coefficient, r was set to 0.5 following the Cochrane Handbook to ensure consistency in effect-size calculation. Because this assumption may slightly change variance estimates in some studies, sensitivity analysis was used to assess the robustness of the overall result.

2.6 Risk-of-bias Assessment

Methodological quality was assessed using the Cochrane Risk of Bias Tool for randomized trials (RoB 1.0). Assessment domains included random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting and other bias. Each domain was rated as low risk, unclear risk or high risk. Because foreign language learning is a behavioral intervention, blinding participants and instructors is usually difficult; this risk was therefore emphasized in interpreting the results. Two reviewers independently completed the assessment, with disagreements resolved through discussion or third-party adjudication.

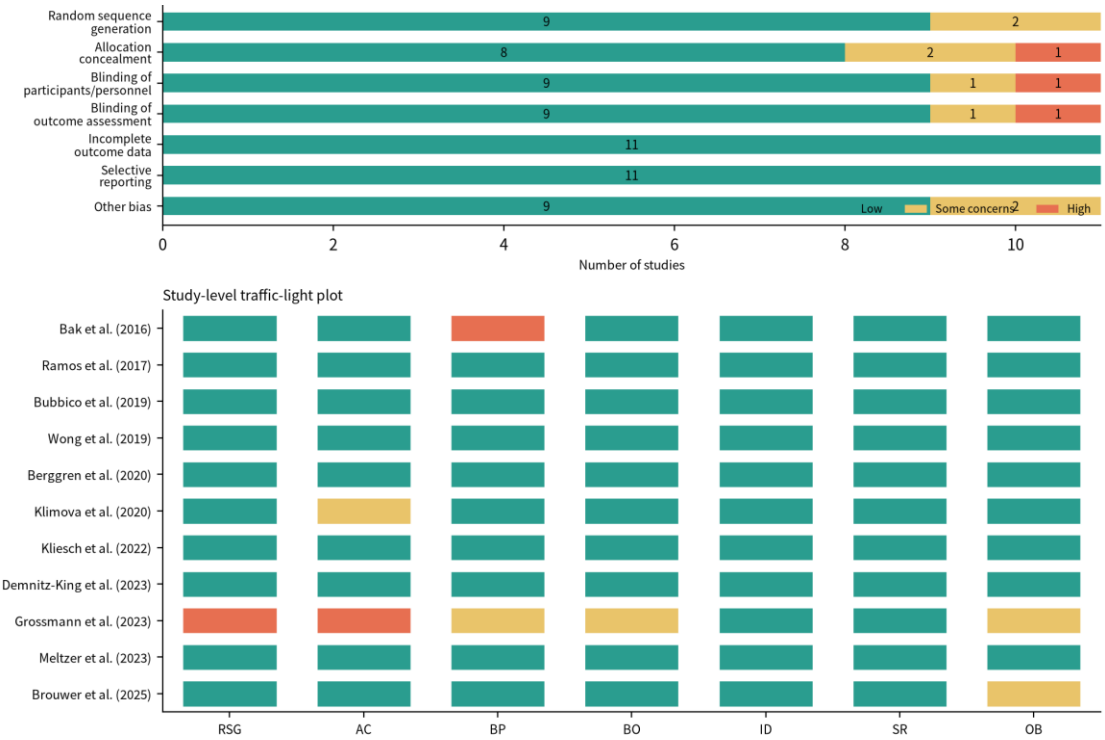


Figure 1. Risk-of-bias assessment of included studies (Cochrane RoB 1.0).

Overall, the methodological quality of included studies varied by domain. Six studies were rated as low risk for random sequence generation, four as unclear and one as high risk. For allocation concealment, only three studies were rated as low risk, and reporting was insufficient in most studies, suggesting that selection bias could affect some effect estimates. Because of the inherent nature of behavioral interventions, 10 studies were rated as high risk for blinding of participants and personnel. If participants in language-learning groups had stronger expectations of benefit, observed cognitive gains may have been amplified through greater engagement, test motivation. Six studies were rated as low risk for blinded outcome assessment, indicating that some studies attempted to reduce detection bias. All studies were rated as low risk for incomplete outcome data and selective reporting, suggesting that attrition and selective reporting were less likely to distort the pooled result. Accordingly, high risk related to blinding was treated as the main methodological limitation, whereas randomization, outcome completeness and reporting were relatively stronger sources of evidence.

2.7 Statistical Analysis

The primary analysis of this review estimated the overall effect of foreign language learning interventions on cognitive function by pooling all eligible objective cognitive outcomes across studies. This approach follows standard practice in behavioral intervention meta-analyses: it provides a comprehensive estimate of global intervention benefit, and maximizes statistical power given the limited number of included studies. It does not assume all cognitive measures tap a single latent construct, but rather summarizes the breadth of cognitive effects across outcomes.

Secondary domain-specific analyses examined four pre-specified cognitive domains: global cognition, executive function, attention, and memory. Quantitative pooling was performed only for domains with three or more independent study estimates; domains with fewer data points were synthesized narratively.

Analyses were conducted in R 4.2.3 using the meta and metaphor packages (Balduzzi et al., 2019). Hedges' g was used as the pooled effect size, corrected for small-sample bias, and reported with 95% CIs. Because included studies differed in populations, intervention protocols, control conditions and cognitive measures, a DerSimonian-Laird random-effects model was used to estimate τ^2 and pool effects. This model assumes that true effects vary across studies rather than that all studies share a single true effect, which is appropriate for meta-analyses of psychological and behavioral interventions with heterogeneous designs. Effect-size interpretation followed Cochrane Handbook conventions: g around 0.20 was considered small, 0.20-0.49 small to moderate, 0.50-0.79 moderate, and 0.80 or greater large (Higgins & Green, 2011).

Between-study heterogeneity was evaluated using Cochran's Q test and I^2. I^2 values of approximately 25%, 50% and 75% were interpreted as low, moderate and high heterogeneity. Leave-one-out sensitivity analysis was used to test robustness. If the direction and statistical significance of the pooled effect remained consistent after removing each study in turn, the overall result was considered relatively robust. Publication bias was explored visually using funnel-plot symmetry. Because Egger regression usually requires a sufficient number of studies for stable power, no strong inference was made when the number of included studies was limited.

Prespecified subgroup analyses used mixed-effects models and Q_between statistics to test between-group differences. This approach allows random variation within subgroups while testing whether average effects differ across subgroups, making it suitable for exploring potential associations between intervention design and effect size. Candidate moderators included participant-level variables (sex distribution, age group, mean years of education and education level), intervention-level variables (L1-L2 linguistic distance, total intervention duration, training frequency, teacher-led versus non-teacher-led format, delivery modality, self-study frequency and inclusion of fMRI measurement), and global cognition measurement type (brief screening versus comprehensive battery). All tests were two-sided, with alpha = 0.05. Results with $p < 0.05$ were considered statistically significant, results approaching but not crossing that threshold were described as borderline, and all others as non-significant.

3. Results

3.1 Literature Screening

The screening process is shown in Figure 2. After title and abstract screening and full-text review, 11 RCTs were included, comprising 840 participants (366 in intervention groups and 474 in control groups). This evidence base indicates that RCT research on late-life foreign language learning remains at an early accumulation stage, so interpretation of the results must consider both pooled effect sizes and the limited number of studies.

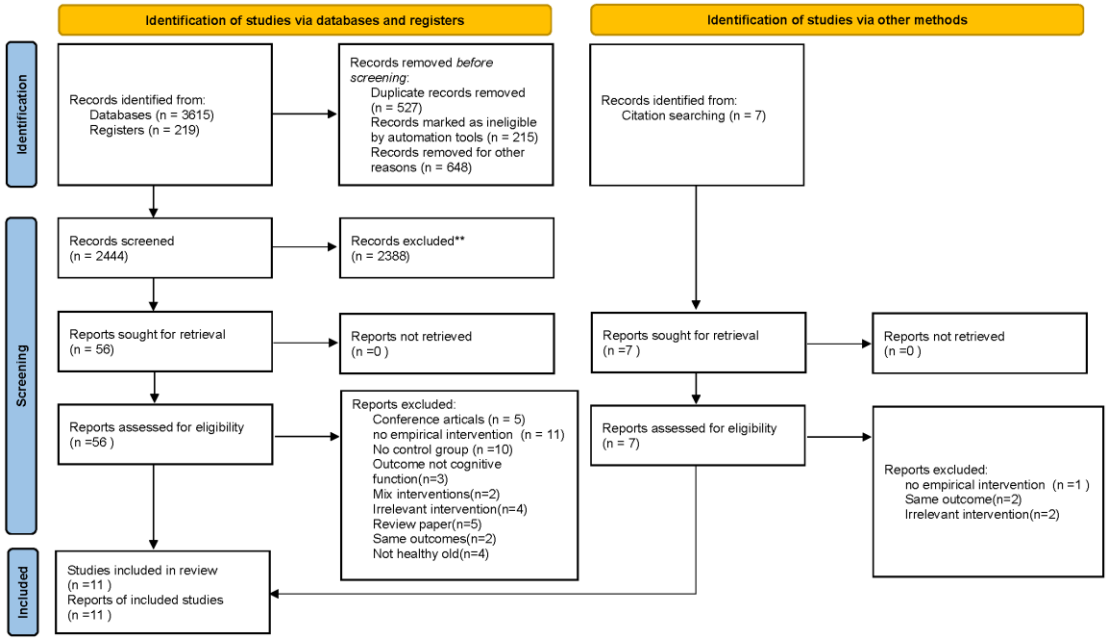


Figure 2. Literature screeningflow diagram.

3.2 Characteristics of Included Studies

Included studies were published between 2016 and 2025, with seven published in 2019 or later, indicating recent growth in the field. Sample sizes ranged from 26 to 153 participants, and the total sample comprised 840 mostly healthy community-dwelling older adults aged 65-75 years. L1-L2 pairings were diverse, including both linguistically close and cross-family combinations. Total intervention duration ranged from one week to 18 months, and training frequency and delivery format varied substantially. These features suggest that the evidence base has some representativeness but also considerable between-study heterogeneity. Detailed study characteristics are shown in Table 1.

Table 1. Characteristics of included studies.

Study	N (Experimental/Control)	L1-L2 Pair	L2 Proficiency at Baseline	Intervention Type	Intervention Duration & Frequency	Cognitive Domains Assessed	Outcome Measures	Effect Size (g)
Brouwer et al. (2025)	15 / 28 (13 Music + 15 Lecture)	Dutch - English	None, Beginner, Intermediate (≤B2 CEFR)	Integrated (focus on listening & speaking)	3 months, biweekly (1.5h/session) + 45min/day self-study	Global cognition, Executive function, Episodic memory, Working memory, Attention/Processing speed, Verbal fluency	MoCA, CSST (mixing/switching cost), Digit Span (forward/backward), Letter-number sequencing, TMT-A/B, VAT-E (paired/free recall), VF, DSST, PPVT	- 0.094
Meltzer et al. (2023)	28 / 48 (24 BrainHQ + 24 No intervention)	English - Spanish	None	Integrated (online app)	16 weeks, 5 days/week (30min/session) + 30min/day self-study	Executive function, Working memory, Processing speed	MoCA, LSBQ	0.640

Study	N (Experimental/Control)	L1-L2 Pair	L2 Proficiency at Baseline	Intervention Type	Intervention Duration & Frequency	Cognitive Domains Assessed	Outcome Measures	Effect Size (g)
Grossmann et al. (2023)	16 / 18 (Waiting list)	German - Spanish	Beginner	Integrated	3 weeks, 5 days/week (1.5h/session)	Executive function, Executive attention, Attention, Verbal fluency, Memory	STROOP (naming/reading interference), INHIB commission errors, Digit Span (forward/backward), NBV S1/S3, WIWO S2/S4, SWITCH switching speed, TMT A/B-A, WAFG/WAFA reaction time	0.679
Demnitz-King et al. (2023)	45 / 90 (45 Meditation + 45 No intervention)	French - English	None/Beginner	Integrated	18 months, 1 time/week (2h/session) + ≥20min/day self-study	Global cognition, Episodic memory, Executive function, Attention	PACC5, Stroop interference, Digit Span backward, TMT-B, Letter fluency, CVLT-II (trials 1-5, immediate/delayed recall), Logical Memory B (immediate/delayed recall)	-0.150
Kliesch et al. (2022)	28 / 33 (17 ACTV + 16 PASV)	German/Swiss German - Spanish	None/Beginner	Integrated	30 weeks, 1 time/week (2h/session) + 2-3h/week self-study	Working memory, Attention (divided/alertness), Verbal fluency, Processing speed		0.523
Klimova et al. (2020)	20 / 22 (No intervention)	Czech - English	Beginner, Lower intermediate	Integrated	3 months, 3 lessons/week (45min/lesson)	Executive function, Visual-spatial, Short-term memory, Attention, Language orientation	MoCA	-0.160
Berggren et al. (2020)	90 / 70 (Relaxation)	Swedish - Italian	None/Very poor	Integrated	11 weeks, 2/week (2.5h/session) + weekly vocabulary self-study	Verbal intelligence, Spatial intelligence, Working memory, Item memory, Associative memory	WASI, N-back, Numerical updating, Face-name, Picture-picture, Word-word, Raven's, Analogies, Syllogisms, Verbal inference	0.433
Wong et al. (2019)	51 / 102 (53 Cognitive games + 49 Music appreciation)	Cantonese/Chinese - English	None	Integrated	6 months, 2-3 days/week (1.5-2.5h/day) + 5h/week self-study	Global cognition, Working memory, Attention, Executive inhibition, Memory, Language, Orientation	ADAS-Cog, Simon response time, Auditory Reading Span, Digit Span (forward/backward), ANT Accuracy	0.821
Bubbico et al. (2019)	14 / 12 (No intervention)	Italian - English	Beginner/None	Integrated	4 months, 1 session/week (2h/session) + 30min/week homework	Global cognition, Executive function, Memory (short/long-term), Attention, Language, Processing speed	MMSE, FAB, TMT A/B/B-A, FAS, IR, DR	-0.339
Ramos et al. (2017)	26 / 17 (No language training)	Spanish - Basque	None	Integrated	8 months, 3 sessions/week (5.5h/week total)	Executive function (task-switching)	SWITCH TRIALS, SWITCH COST, NON-SWITCH TRIALS	0.212
Bak et al. (2016)	33 / 34 (16 Active + 18 Passive)	English - Scottish Gaelic	Beginner, Elementary, Intermediate	Integrated	1 week, daily (intensive)	Attention (sustained, selective/inhibition, switching)	ETD, ETR, ET	-0.150

3.3 Primary Analysis: Overall Cognitive Effect

This primary analysis pooled all eligible cognitive outcomes across 11 RCTs (840 participants) to estimate the overall effect of foreign language learning relative to control conditions. The overall effect represents a broad cognitive performance index rather than a unitary neuropsychological construct. The random-effects model yielded a small-to-moderate, statistically significant pooled estimate of overall cognitive performance among healthy older adults (Hedges' g = 0.309, 95% CI: 0.153- 0.465, Z=3.875, $p<0.001$). Heterogeneity was moderate (Cochran's Q = 22.323, df = 10, p = 0.014; I^2 approximately 55.2%). Individual study effect sizes ranged from small negative to large positive values (g =-0.339 to 0.821). Most studies favored the intervention group, but magnitude varied substantially. Seven studies

had 95% CIs that did not cross zero. Because this primary estimate pools heterogeneous cognitive outcomes, it summarizes the breadth of cognitive effects rather than any single construct and should be interpreted cautiously.

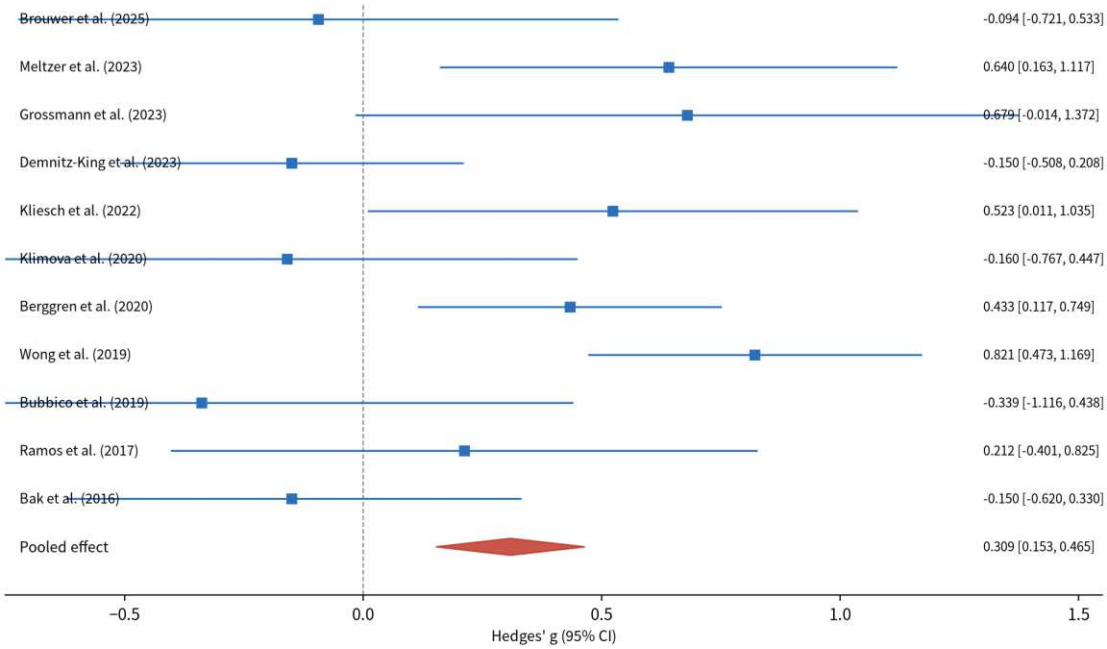


Figure 3. Forest plot of the effect of foreign language learning on overall cognitive performances in healthy older adults.

Forest-plot distributions showed that individual study effects ranged from negative to large positive effects (-0.339 to 0.821). Most studies favored the intervention, but effect magnitudes were inconsistent. Seven studies had 95% CIs that did not include zero. Wong et al. (2019) reported the largest effect (g = 0.821), whereas Bubbico et al. (2019) reported the smallest and negative effect (g = -0.339), indicating the need to consider differences in intervention design and methodology.

3.4 Effects Across Cognitive Domains

Consistent with the prespecified analysis plan, domain-specific analyses were treated as secondary and exploratory. Global cognition was the only domain with enough studies to support quantitative pooling. The pooled effect across seven studies was g=0.292 (95% CI: 0.116-0.469, $p<0.001$), a small-to- moderate effect, although heterogeneity was high (Cochran's Q=19.397, df=6, p=0.004; I^2 approximately 69. 1%). When stratified by measurement type, brief screening tools such

as MMSE and MoCA (k = 5) yielded a pooled effect of g = 0.258 (95% CI: 0.034-0.483, p= 0.024), whereas comprehensive cognitive batteries (k = 2) yielded g = 0.349 (95% CI: 0.061 to 0.636, p= 0.018). The difference between these subgroups was not statistically significant (Q_between = 0.236, df = 1, p = 0.627), and current data did not show that measurement type significantly altered global cognitive benefit.

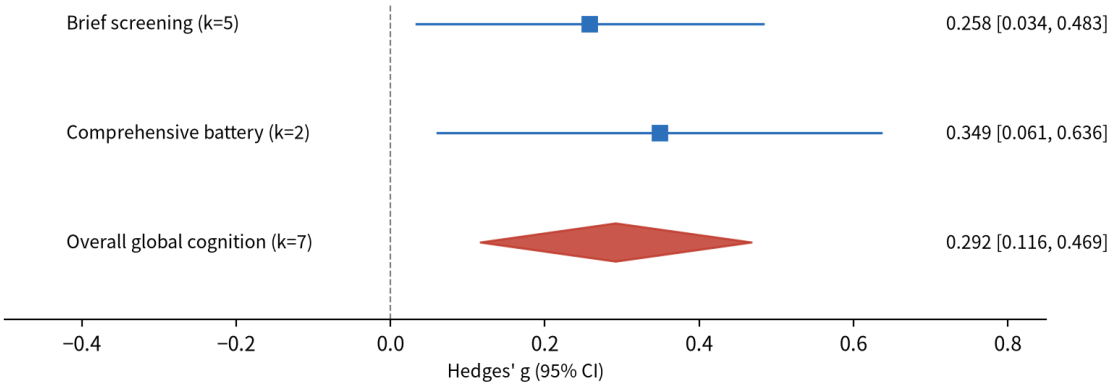


Figure 4. Global cognition effects and stratified analysis by measurement type.

Other cognitive domains (executive function, working memory and attention) included fewer than three contributing studies per standardized outcome measure, precluding independent quantitative pooling. To fully address the second research question, a narrative synthesis was conducted for these domains. For executive function, two smaller studies reported positive effects: Grossmann et al. (2023) found improved response inhibition among participants with lower baseline cognition following three weeks of intensive Spanish training, and Ramos et al. (2017) observed advantages in task-switching performance after eight months of Basque learning. However, three larger trials (Berggren et al., 2020; Brouwer et al., 2025; Demnitz-King et al., 2023) found no statistically significant between-group differences in executive function composite scores or individual subdomains such as inhibitory control and cognitive flexibility. For working memory, Wong et al. (2019) and Kliesch et al. (2022) reported descriptive improvements in digit span performance and resting-state working memory network characteristics, but Berggren et al. (2020) detected no working memory advantage of language learning relative to an active relaxation control in a larger, well-powered sample. For attention, findings were similarly mixed: Bak et al. (2016) found no sustained or selective attention benefit of one-week intensive Gaelic training, while several

other studies recorded weak, non-significant trends favouring the intervention group on attention subscales. Overall, individual trial findings suggest potential domain-specific benefits in select executive function and working memory outcomes, but the evidence base is inconsistent and underpowered. Domain-specific conclusions should therefore be interpreted as exploratory rather than definitive.

3.5 Heterogeneity, Sensitivity and Publication-bias Analyses

Overall analyses showed moderate heterogeneity, and the global cognition subgroup showed higher heterogeneity, suggesting real variation in effects across studies. Leave-one-out sensitivity analysis showed that the pooled effect fluctuated between 0.267 and 0.378 after removing individual studies, while the direction remained positive and statistical significance was unchanged. Thus, the overall result was not overly driven by any single study. Visual inspection of the funnel plot suggested approximate symmetry, but with only 11 included studies, Egger regression had insufficient power and any inference about publication bias should be considered preliminary.

In addition to the leave-one-out analysis, a data-format sensitivity analysis was conducted to test whether the overall result was driven by studies requiring statistical conversion. We re-ran the primary random-effects model using only studies that directly reported pre- and post- intervention means and standard deviations, excluding studies requiring median-to-mean conversion or posttest-only estimation.

This restricted analysis included 7 studies and 596 participants. The pooled effect remained positive and statistically significant (Hedges' g = 0.287, 95% CI: 0.114-0.460, p = 0.001), with moderate heterogeneity (I^2= 48%). The effect magnitude fell within the range observed in the leave-one-out analysis, and the direction and significance were unchanged. This result supports the robustness of the primary pooled estimate and indicates that the overall conclusion is not dependent on converted or imputed data. Visual inspection of the funnel plot suggested approximate symmetry. Formal Egger regression was not interpreted as robust due to the small number of included studies (k = 11). Full results of all sensitivity analyses are summarized in Supplementary Table S3 (Sensitivity Analyses)

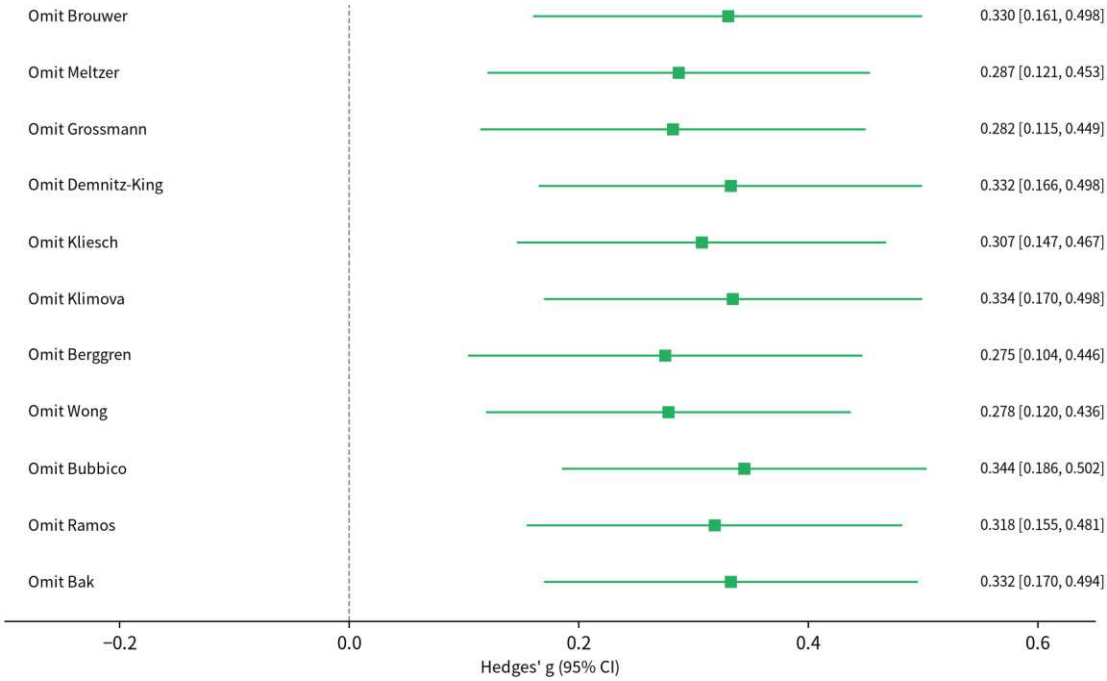


Figure 5. Leave-one-out sensitivity analysis.

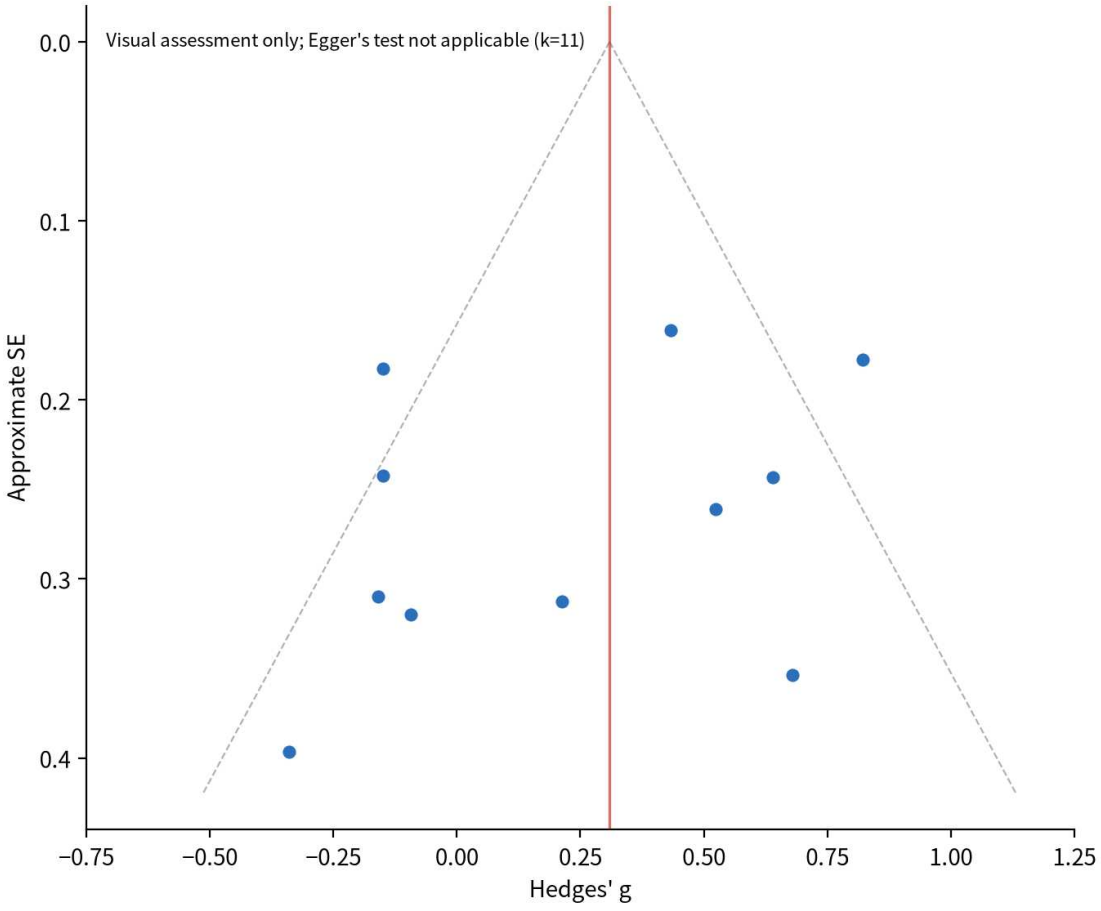


Figure 6. Funnel plot for the overall cognitive effect.

3.6 Moderator Subgroup Analyses

To explore potential sources of between-study heterogeneity, 12 prespecified candidate moderators were examined using mixed-effects subgroup analyses in an exploratory manner. Given the limited number of included studies and small sample sizes in several subgroups, all subgroup findings should be interpreted as preliminary observational associations rather than definitive causal moderators. Moreover, the online, self-directed and cross-family subgroups each contained only two to three studies and were partly overlapping, so these differences should be regarded as hypothesis-generating rather than confirmatory. Four intervention-level variables were statistically associated with larger observed effect sizes: linguistic distance, training frequency, delivery modality and teacher-led versus self-directed format. Subgroup forest plots for each moderator analysis are provided in Supplementary Figures S1-S4. No significant associations were identified for participant-level variables. Detailed subgroup results are shown in Table 2.

Table 2. Subgroup analyses of moderator variables

Characteristics	Studies (n)	Effect size and 95% CI				Heterogeneity			
		Hedges's g	Lower limit	Upper limit	*p* value	Q	df(Q)	*p* value	I^2
Female Percentage						2.079	2	0.354	
male dominant	2	0.373	-0.046	0.791	0.081				
balanced	6	0.266	0.065	0.467	0.010				
female dominant	2	0.564	0.210	0.918	0.002				
Overall	10	0.343	0.182	0.505	0.000	19.762	9	0.019	
Age Classification						0.054	1	0.817	
younger elderly	6	0.292	0.083	0.502	0.006	8.929	5	0.112	
integrated	5	0.330	0.096	0.563	0.006	13.341	4	0.010	
Overall	11	0.309	0.153	0.465	0.000	22.323	10	0.014	
Mean Years of Education						2.143	1	0.143	
high	7	0.240	0.054	0.425	0.011	10.266	6	0.114	
moderate-high	1	-0.339	-1.091	0.413	0.377	0.000	0	1.000	
Overall	8	0.207	0.026	0.387	0.025	12.409	7	0.088	
Education Level						1.228	1	0.268	
high	3	0.242	-0.148	0.632	0.224	3.613	2	0.164	
moderate-high	1	-0.160	-0.755	0.435	0.598	0.000	0	1.000	
Overall	4	0.121	-0.205	0.448	0.466	4.841	3	0.184	
L1–L2 linguistic distance						4.630	1	0.031	

Characteristics	Studies (n)	Effect size and 95% CI				Heterogeneity			
		Hedges's g	Lower limit	Upper limit	*p* value	Q	df(Q)	*p* value	I^2
same	9	0.219	0.043	0.395	0.015	14.972	8	0.060	
different families	2	0.633	0.299	0.968	0.000	2.721	1	0.099	
Overall	11	0.309	0.153	0.465	0.000	22.323	10	0.014	
Total Intervention Duration						0.032	1	0.857	
long	9	0.305	0.142	0.468	0.000	20.046	8	0.010	
short	2	0.355	-0.173	0.884	0.188	2.245	1	0.134	
Overall	11	0.309	0.153	0.465	0.000	22.323	10	0.014	
Duration Frequency						7.196	2	0.027	
low	4	-0.010	-0.291	0.271	0.943	4.449	3	0.217	
moderate	4	0.439	0.225	0.652	0.000	7.916	3	0.048	
High	3	0.488	0.101	0.874	0.013	2.763	2	0.251	
Overall	11	0.309	0.153	0.465	0.000	22.323	10	0.014	
With Teacher or not						9.425	1	0.002	
yes	9	0.177	0.000	0.354	0.050	12.637	8	0.125	
no	2	0.761	0.433	1.088	0.000	0.261	1	0.609	
Overall	11	0.309	0.153	0.465	0.000	22.323	10	0.014	
Modality						6.660	2	0.036	
online	3	0.606	0.309	0.902	0.000	4.993	2	0.082	
offline	7	0.160	-0.033	0.353	0.103	10.670	6	0.099	
offline+online	1	0.523	-0.059	1.106	0.078	0.000	0	1.000	

Characteristics	Studies (n)	Effect size and 95% CI				Heterogeneity			
		Hedges's g	Lower limit	Upper limit	*p* value	Q	df(Q)	*p* value	I^2
Overall	11	0.309	0.153	0.465	0.000	22.323	10	0.014	
Self-Study Frequency						1.095	1	0.295	
high	7	0.353	0.177	0.530	0.000	17.321	6	0.008	
not clear	4	0.153	-0.177	0.484	0.363	3.908	3	0.272	
Overall	11	0.309	0.153	0.465	0.000	22.323	10	0.014	
fMRI or not						2.977	1	0.084	
yes	1	-0.339	-1.091	0.413	0.377	0.000	0	1.000	
no	10	0.338	0.179	0.497	0.000	19.347	9	0.022	
Overall	11	0.309	0.153	0.465	0.000	22.323	10	0.014	
global cognition						0.236	1	0.627	
Brief screening global cognition	5	0.258	0.034	0.483	0.024	8.237	4	0.083	
Comprehensive global cognition	2	0.349	0.061	0.636	0.018	10.924	1	0.001	
Overall	7	0.292	0.116	0.469	0.000	19.397	6	0.004	

3.6.1 Linguistic Distance

L1-L2 linguistic distance was associated with larger observed effect sizes in this exploratory analysis. When studies were grouped by whether L1 and L2 belonged to the same language family, the same-family subgroup included nine studies, such as German-Spanish and Dutch- English, with a pooled effect of g= 0.219 (95% CI: 0.043 to 0.395, p = 0.015). The cross-family subgroup included two studies, Cantonese/Chinese-English and Spanish-Basque, with a pooled effect of g = 0.633 (95% CI: 0.299 to 0.968, p < 0.001). The between-group difference was statistically significant (Q_between = 4.630, p = 0.031), suggesting that linguistic distance may moderate cognitive effects. The same-family subgroup mostly showed small effects, whereas the cross-family subgroup showed larger effects, possibly reflecting greater phonological, lexical and syntactic differentiation demands. Because the cross-family subgroup contained only two studies, however, this finding should be treated as exploratory and should not be overgeneralized; the two cross-family studies also differed in delivery format and sample characteristics, so the association is likely confounded and requires dedicated experimental testing.

3.6.2 Training Frequency

Moderate- and high-frequency training were associated with statistically significant positive effects, whereas low-frequency training was not. When weekly training was categorized as low (<3 h/week), moderate (3-6 h/week) or high (>6 h/week), the low-frequency subgroup included four studies and showed no significant effect (g = -0.010, 95% CI: -0.291 to 0.271, p = 0.943). The moderate-frequency subgroup included four studies and showed g = 0.439 (95% CI: 0.225 to 0.652, p < 0.001), whereas the high-frequency subgroup included three studies and showed g = 0.488 (95% CI: 0.101 to 0.874, p = 0.013). Differences across the three subgroups were statistically significant (Q_between = 7.196, p = 0.027).

The low-frequency subgroup had an effect close to zero, whereas moderate- and high-frequency groups showed positive effects. This pattern suggests that foreign language learning may require a minimum weekly training threshold to operate as a cognitive intervention, and that measurable benefits may be easier to detect when weekly training reaches approximately three

hours or more. Because only three studies were included in the high-frequency subgroup, further evidence is needed to confirm stability.

3.6.3 Teacher-led Versus Self-directed Learning

The self-directed online learning subgroup showed larger effects than the teacher-led intervention subgroup. Nine teacher-led studies yielded a pooled effect of g = 0.177 (95% CI: 0.000 to 0.354, p = 0.050), a borderline result. Two self-directed online studies yielded g = 0.761 (95% CI: 0.433 to 1.088, p < 0.001). The between-group difference was statistically significant (Q_between = 9.425, p = 0.002).

Teacher-led interventions showed smaller effects overall, whereas self-directed online learning showed larger effects. This may relate to the additional executive demands of self- pacing, immediate feedback, progress monitoring and strategy adjustment. Because this subgroup contained only two self-directed studies that also delivered training online, the result is likely confounded with delivery modality and learner self-selection; it should be treated as hypothesis-generating.

3.6.4 Delivery Modality

Fully online delivery showed a statistically significant moderate effect, whereas fully offline delivery was not statistically significant. Three online studies yielded g = 0.606 (95% CI: 0.309 to 0.902, p < 0.001). Seven offline studies yielded g = 0.160 (95% CI: -0.033 to 0.353, p = 0. 103). Only one hybrid online-offline study was available, with g = 0.523 (95% CI: -0.059 to 1.106, p = 0.078). Differences across modalities were statistically significant (Q_between = 6.660, p = 0.036). Only three studies were fully online, and they overlapped with the self-directed and cross-family subgroups, so this difference should be interpreted cautiously.

3.6.5 Other Moderators

Sex distribution, age classification, mean years of education, education level, total intervention duration, self-study frequency and inclusion of fMRI measurement did not show statistically significant between-group differences (all p > 0.05). This does not mean that demographic factors are irrelevant, but rather that within the available sample range, differences in effect size may be driven more strongly by course factors that can be designed and adjusted. Non-significant moderation by

age, sex and education may reflect the relative homogeneity of healthy older-adult samples or limited statistical power in subgroup analyses.

4. Discussion

This systematic review and meta-analysis synthesized randomized controlled trial evidence from 11 trials involving 840 cognitively healthy older adults to quantify cognitive changes induced by structured foreign language learning. The pooled random effects estimate was a statistically significant small-to-moderate estimate of overall cognitive performance, with a pooled Hedges' g of 0.309 (95 % CI: 0.153-0.465, $p < 0.001$) relative to diverse control conditions. This estimate remained stable across multiple sensitivity analyses, including leave-one-out procedures and restricted re-analysis limited to studies reporting raw pre-post means and standard deviations, which yielded highly consistent effect-size magnitudes. Moderate between-study heterogeneity ($I^2 = 55.2\%$) indicated that cognitive gains were not universal across different implementation contexts, meaning structured foreign-language learning functions as a conditional cognitive-enrichment activity rather than an intervention that guarantees cognitive improvement for all older adult learners. When interpreting this pooled global effect size, it is important to note that this composite estimate aggregates heterogeneous cognitive outcome measures across trials and reflects broad functional cognitive change, rather than change within a single unitary neuropsychological construct.

The observed small-to-moderate global cognitive effect offers nuanced empirical input for existing theories of cognitive ageing. From the perspective of cognitive reserve frameworks, this result supports the proposition that sustained intellectually demanding activities can drive measurable cognitive adaptation in later life (Stern, 2009), yet it also qualifies earlier observational bilingual advantage accounts that expected large, automatic cognitive gains from language related mental engagement. The significant pooled behavioral effect aligns with predictions derived from the Scaffolding Theory of Aging and Cognition and the Dynamic Restructuring Model, which posit that the ageing brain retains neuroplastic capacity and can form compensatory neural networks in response to novel complex learning tasks (Park & Reuter-Lorenz, 2009; Pliatsikas, 2020). At the same time, the moderate

magnitude of the effect and substantial cross-study heterogeneity partially counter strong versions of the bilingual-advantage hypothesis extrapolated from observational cohorts, while language based cognitive stimulation can produce measurable behavioral benefits, such benefits are not large and are highly dependent on how learning interventions are designed and delivered. One plausible mechanistic explanation for the global cognitive improvement is that structured foreign language learning repeatedly engages attentional control, error monitoring, rule extraction, and working memory updating simultaneously. This multi-component cognitive load creates recurring opportunities for cognitive adaptation, consistent with how cognitive reserve theory conceptualizes enrichment-driven brain change (Hertzog et al., 2008). Such mechanism-related interpretations remain hypotheses, however; they cannot be verified within a meta-analysis and require primary studies designed to test mediating processes. The domain specific secondary analyses further refine this picture. Among individual cognitive subdomains, only global cognition yielded sufficient study numbers for quantitative pooling, returning a small-to-moderate significant pooled effect ($g = 0.292$). By contrast, effects for executive function, working memory, and attention were mixed and could not be reliably aggregated. This pattern suggests that cognitive-related changes induced by foreign language learning manifest as broad-based functional improvement rather than selective far transfer to isolated executive or memory subconstructs. These results offer a provisional reconciliation of seemingly contradictory prior RCT evidence: some individual trials reported large positive cognitive shifts (Wong et al., 2019), while others found null effects (Berggren et al., 2020). Instead of treating these conflicting results as mutually exclusive, the present meta-analysis indicates that positive cognitive change can occur under appropriate intervention conditions, yet may be absent when implementation parameters are sub-optimal. Measurement tool type did not significantly moderate global cognition effect sizes in subgroup comparisons, implying that the observed global cognitive benefit is not merely an artefact arising from selection of brief screening versus comprehensive neuropsychological test batteries.

Among the 12 prespecified moderators examined in exploratory subgroup analyses, four intervention-level variables emerged as statistically significant predictors of effect-size magnitude:

L1-L2 linguistic distance, weekly training frequency, self-directed versus teacher-led instructional format, and intervention delivery modality.

Linguistic distance between the native and target second language showed a significant moderating influence, whereby languages drawn from different language families produced substantially larger cognitive effect sizes than language pairs within the same language family. This pattern can be understood through neuroplasticity oriented reasoning embedded within the Dynamic Restructuring Model, greater phonological, lexical, and syntactic dissimilarity increases the total cognitive processing burden during language acquisition, placing higher recurring demands on executive and mnemonic systems and thereby providing stronger stimulation for brain compensatory reorganization (Pliatsikas, 2020). This empirical result is partially consistent with theoretical expectations from Carthery-Goulart et al. (2023), who argued that higher task complexity may amplify cognitive enrichment outcomes. Importantly, this subgroup finding runs counter to the competing concern that extreme linguistic dissimilarity would trigger excessive learner frustration, reduce adherence, and erase cognitive benefits (Ware et al., 2021), at least within the sample of RCTs synthesized here. Nevertheless, the cross-family subgroup contained only two independent studies, so this observed association cannot be interpreted as definitive causal evidence. It remains possible that unmeasured co-occurring factors such as course material design or heightened participant motivation partly explain the larger effect sizes observed for cross-family language learning.

Moderate-to-high weekly practice yielding significant cognitive improvements whereas low frequency training produced effects near zero. This outcome aligns well with cognitive reserve informed dose-response assumptions from prior cognitive training meta-analyses (Lampit et al., 2014). Cognitive reserve theory emphasises that cognitive enrichment gains stem from repeated, sustained mental engagement; sporadic low-dose practice may fail to supply sufficient recurring cognitive challenge to consolidate new linguistic representations and trigger the neuroplastic adaptations required to translate learning activity into measurable behavioural cognitive change (Stern, 2009). The present dataset did not show clear evidence for cognitive fatigue blunting benefits among high frequency subgroups, a risk highlighted in previous theoretical work (Hou et al., 2020). Even so, real world

563 community programmes still need to balance weekly training dosage against older adults' time
564 constraints and subjective fatigue. These subgroup results suggest there exists a minimum weekly
565 exposure threshold for foreign language learning to operate as an effective cognitive intervention,
566 reinforcing that simply enrolling in language classes without adequate regular practice is unlikely to
567 deliver detectable cognitive benefits.

568 Self-directed learning formats were associated with considerably larger cognitive effect sizes
569 compared with teacher-led instruction. This difference may originate from the elevated self-regulatory
570 cognitive demands intrinsic to self-directed study. Self-directed learning compels learners to
571 independently schedule practice, monitor performance, detect errors, and adjust learning strategies,
572 continuously exercising metacognition, cognitive flexibility, and inhibitory control beyond the work of
573 acquiring second-language vocabulary and grammar (Knowles et al., 2015). Within the interpretive
574 lens of the Scaffolding Theory of Aging and Cognition, these recurring self-regulatory demands impose
575 extra cognitive challenge that invites additional recruitment of compensatory brain networks (Park &
576 Reuter-Lorenz, 2009). The finding partially supports hypotheses raised by Reuter-Lorenz and Park
577 (2014) regarding the potential cognitive value of autonomous learning activities for older adults. It
578 should be noted, however, that self-directed protocols in existing RCTs frequently co-occurred with
579 online digital platforms and higher intrinsic learner motivation, creating potential confounding.
580 Teacher-led interventions still returned a borderline positive pooled effect, demonstrating that they are
581 not cognitively inert; they simply produced smaller observable cognitive gains under existing trial
582 conditions.

583 Fully online delivery yielded significant moderate cognitive benefits, while offline classroom
584 based interventions showed non-significant effect sizes. Digital online learning environments
585 commonly provide instant automated feedback, adaptive difficulty adjustments, and frequent real-time
586 progress monitoring, features that keep executive control processes activated throughout learning
587 episodes (Chan et al., 2022). Such continuous, structured cognitive stimulation fits well with cognitive
588 enrichment assumptions from cognitive reserve theory. This meta-analytic result lends tentative
589 quantitative support to arguments from Parmaxi and Zaphiris (2017) regarding the cognitive potential

of technology assisted language learning for older populations. On the other hand, the null-to-small effect observed for offline studies appears to contradict viewpoints that prioritize face-to-face social interaction as the primary driver of cognitive benefits in group-based learning programs (Brouwer et al., 2025). One possible reconciliation is that offline courses do supply valuable social enrichment, yet this social component alone may not generate sufficient cognitive processing load to produce the magnitude of behavioral cognitive change seen in online self-directed language interventions in the current evidence base. Hybrid format evidence remained limited to a single study and could not support firm subgroup conclusions.

These meta-analytic findings carry preliminary implications for designing foreign-language learning programmes intended to serve as non-pharmacological cognitive enrichment tools for community dwelling healthy older adults. Programme planners should treat any dosage guidance as preliminary and hypothesis-generating rather than prescriptive; sporadic low-frequency sessions appear insufficient to generate measurable cognitive-related benefits (Lampit et al., 2014). Any specific weekly-volume recommendation (e.g., roughly three hours) would require experimental confirmation and should be balanced against older adults' time constraints and potential fatigue (Hou et al., 2020); it is not established by the present data. Linguistic distance between the learner's native and target language warrants experimental investigation: selecting a target language from a different language family may, under sustained engagement and manageable frustration, yield larger estimates, but this remains a hypothesis (Carthery-Goulart et al., 2023). Online and self-directed learning formats warrant further investigation for cognitive health-oriented interventions (Chan et al., 2022; Parmaxi & Zaphiris, 2017). Nevertheless, self-directed online learning presupposes adequate digital literacy and intrinsic motivation; teacher-supported offline classes remain a valuable alternative for older adults who require more social and instructional scaffolding, even if observed cognitive effect sizes were smaller in the current evidence (Brouwer et al., 2025). From a public health perspective, foreign language learning should be framed as one optional behavioural component of multi-domain healthy ageing strategies, rather than a standalone anti-cognitive-decline tool (Livingston et al., 2024; World Health Organization, 2026). Real world programmes ought to monitor adherence, learning burden and

subjective well-being alongside cognitive outcomes. Cognitive benefits should not be over-promised to older adult participants; expectations should emphasise that any cognitive gains are conditional on implementation features and remain preliminary pending confirmatory trials.

Several important limitations must be acknowledged when interpreting the present findings. The total evidence base remains modest, comprising only 11 RCTs, and several moderator subgroups contained very few studies, increasing statistical uncertainty for subgroup comparisons (Higgins et al., 2024). As a behavioral intervention, complete participant and instructor blinding is practically unachievable; accordingly, participant expectations, learning motivation, and Hawthorne type effects may inflate observed cognitive improvements (Brouwer et al., 2025). Control conditions across included trials were highly heterogeneous, ranging from wait-list and no intervention controls to active comparison groups such as music, meditation and cognitive games. Variation in social contact, time investment and task engagement across control arms complicates direct cross trial comparison of effect magnitudes (Hertzog et al., 2008). Some effect size calculations relied on statistical assumptions such as assumed pre-post correlation coefficients, even though sensitivity analyses indicated main outcomes were not driven by converted data. Almost all trials only reported immediate post-intervention assessments; consequently, we cannot determine whether cognitive benefits persist over months or years, or whether gains disappear once language training ceases (Kliesch et al., 2022). Finally, the review protocol was not prospectively registered, which reduces transparency for assessing selective analytical choices (Page et al., 2021). Additionally, earlier systematic reviews have already addressed foreign-language learning and cognition in older adults; although the present synthesis is, to the best of our knowledge, the first meta-analysis restricted to RCTs in cognitively healthy older adults, readers should nonetheless weigh its incremental contribution against that prior work. Collectively, these constraints mean our conclusions should be regarded as preliminary and require further empirical validation.

Several avenues for future research emerge from the current meta-analysis to move this research field forward. Large scale, multicenter, pre-registered randomized controlled trials with long term follow-up are urgently needed. Existing studies are mostly single-site small-sample trials with only

short-term post-testing. Pre-registration of hypotheses, primary cognitive endpoints and analytical plans will reduce researcher degrees of freedom and outcome-reporting bias (Page et al., 2021). Adding extended follow-up assessments (6-12 months after intervention completion) will clarify whether cognitive gains are transient or durable, and whether sustained language practice is required to maintain observed cognitive benefits (Ngandu et al., 2015). Future trials should adopt standardized cognitive outcome measures and employ rigorously matched active control conditions. Heterogeneous neuropsychological test batteries have hindered cross-trial comparison and stable domain-specific conclusions (Higgins et al., 2024). Consensus on core outcome sets for older adult cognitive enrichment interventions would improve evidence comparability. Critically, active control groups need to be closely matched with language-learning interventions on time commitment, social interaction, participant expectation effects and general mental engagement. Such designs enable researchers to separate true language specific cognitive effects from nonspecific stimulation produced by any intellectually engaging group activity (Hertzog et al., 2008). In addition, intervention-optimization trials are required to directly test candidate moderators and underlying neurocognitive mechanisms. Rather than relying on post-hoc subgroup meta-analysis, primary studies should experimentally manipulate key intervention features identified in this review: linguistic distance, weekly training dosage, online-offline delivery, and self-directed versus teacher-led instruction. Combining behavioral cognitive assessments with neural measures such as fMRI or EEG can help unpack whether cognitive changes arise from executive control engagement, neural compensation, or general mental enrichment (Pliatsikas, 2020; Bubbico et al., 2019). Direct experimental testing of moderators will help establish causal inferences and inform the design of optimized, scalable language-learning interventions for promoting healthy cognitive ageing.

Foreign language learning in later life is more than an educational activity. Under appropriate implementation conditions, it represents a promising behavioral pathway to foster cognitive reserve for community dwelling older adults (Stern, 2009).

5. Conclusion

Drawing on data from 11 eligible randomized controlled trials encompassing 840 cognitively healthy community dwelling older adults, this meta-analysis provides preliminary evidence that structured late-life foreign-language learning is associated with a small-to-moderate estimate of overall cognitive performance, with the most robust pooled estimate observed for global cognition (Antoniou et al., 2013; Wong et al., 2019). These associations were not an inherent by-product of language-learning participation alone but appeared contingent upon intervention-specific features, consistent with prior cognitive enrichment frameworks (Stern, 2009). Among twelve pre-specified moderators, four intervention-level factors showed exploratory associations with effect magnitude; these were based on small, partly overlapping subgroups and should be treated as hypothesis-generating: greater linguistic distance between native and target language, moderate-to-high weekly training dosage, self-directed instructional formats, and fully online delivery were linked to larger cognitive improvements (Carthery-Goulart et al., 2023; Lampit et al., 2014). By contrast, demographic indicators including age, sex distribution, and educational background did not emerge as significant moderators within the available evidence base, implying that baseline cognitive reserve proxies do not reliably predict intervention response in this sample (Valenzuela & Sachdev, 2006). Such patterns tentatively suggest that optimizing programme design, rather than selecting learners by demographic profile, may be the more promising route, pending confirmation in intervention-optimisation trials. In addition, the evidence base was insufficient to determine whether benefits are consistent across individual cognitive domains (executive function, working memory and attention could not be pooled quantitatively), so the second research question remains only partially answered; overall cognitive outcomes remain preliminary for older adults engaging in foreign-language study.

These empirical findings carry concrete practical implications for real-world programme development targeting older adult populations (Livingston et al., 2024; World Health Organization, 2026). For community based cognitive enrichment initiatives aiming to harness foreign language learning, the observed weekly-frequency pattern, though preliminary, warrants direct experimental testing (Lampit et al., 2014). Where learner motivation and self-efficacy permit, online self-directed curricula warrant further study (Chan et al., 2022; Parmaxi & Zaphiris, 2017); for older adults lacking

sufficient digital literacy or requiring social support, teacher-led offline classes remain a viable alternative despite smaller observed cognitive effect sizes (Brouwer et al., 2025). Programme organizers should balance cognitive and stimulation goals against participant burden and frustration when selecting target languages with large linguistic distance (Ware et al., 2021). From a public health standpoint, foreign language learning ought to be positioned as one optional behavioral component within multidomain healthy ageing strategies instead of being marketed as a standalone intervention against cognitive decline. Real-world implementations should routinely track adherence, subjective well-being, and learning load alongside cognitive test outcomes and manage participant expectations by clarifying that any cognitive improvements appear to depend heavily on implementation and remain preliminary.

This body of evidence is constrained by a limited total number of included RCTs and small sample sizes in several moderator subgroups, creating statistical uncertainty for subgroup comparisons (Higgins et al., 2024); future work should conduct large-scale multicenter trials to corroborate the present exploratory subgroup patterns. As a behavioral intervention, complete blinding of participants and instructors cannot be achieved, potentially allowing expectation-driven Hawthorne effects to inflate measured cognitive benefits (Brouwer et al., 2025); subsequent studies should adopt rigorously matched active control conditions to disentangle language specific effects from non-specific social and motivational influences (Hertzog et al., 2008). Heterogeneous control arms and variable neuropsychological test batteries across trials complicate cross study comparisons (Higgins et al., 2024); consensus core outcome sets for older adult cognitive enrichment interventions would improve evidence comparability. Nearly all available evidence captures only immediate post-intervention outcomes, leaving long term durability of cognitive gains unknown (Kliesch et al., 2022); extended 6-12-month follow-up assessments are needed to test whether benefits persist after training ceases (Ngandu et al., 2015). Finally, the present review protocol lacked prospective pre-registration; future meta-analyses should pre-register hypotheses, inclusion criteria and analytical workflows to minimize selective analytical bias (Page et al., 2021).

References

[1] Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*(2), 207–245. https://doi.org/10.3102/0034654310368803

[2] Alzheimer's Association. (2024). 2024 Alzheimer's disease facts and figures. *Alzheimer's & Dementia, 20*(5), 3708–3821. https://doi.org/10.1002/alz.13809

[3] Antoniou, M., Gunasekera, G. M., & Wong, P. C. M. (2013). Foreign language training as cognitive therapy for age-related cognitive decline: A hypothesis for future research. *Neuroscience & Biobehavioral Reviews, 37*(10), 2689–2698. https://doi.org/10.1016/j.neubiorev.2013.09.004

[4] Bak, T. H., Long, M. R., Vega-Mendoza, M., & Sorace, A. (2016). Novelty, challenge, and practice: The impact of intensive language learning on attentional functions. *PLOS ONE, 11*(4), e0153485. https://doi.org/10.1371/journal.pone.0153485

[5] Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health, 22*(4), 153–160. https://doi.org/10.1136/ebmental-2019-300117

[6] Berggren, R., Nilsson, J., Brehmer, Y., Schmiedek, F., & Lövdén, M. (2020). Foreign language learning in older age does not improve memory or intelligence: Evidence from a randomized controlled study. *Psychology and Aging, 35*(2), 212–219. https://doi.org/10.1037/pag0000439

[7] Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences, 16*(4), 240–250. https://doi.org/10.1016/j.tics.2012.03.001

[8] Brouwer, J., van den Berg, F., Knooihuizen, R., Loerts, H., & Keijzer, M. (2025). The effects of language learning on cognitive functioning and psychosocial well-being in cognitively healthy older adults: A semi-blind randomized controlled trial. *Aging, Neuropsychology, and Cognition, 32*(2), 270–306. https://doi.org/10.1080/13825585.2024.2384107

[9] Bubbico, G., Chiacchiaretta, P., Parenti, M., Di Marco, M., Panara, V., Sepede, G., Ferretti, A., & Perrucci, M. G. (2019). Effects of second language learning on the plastic aging brain: Functional connectivity, cognitive decline, and reorganization. *Frontiers in Neuroscience, 13*, 423. https://doi.org/10.3389/fnins.2019.00423

[10] Buckner, R. L. (2004). Memory and executive function in aging and AD: Multiple factors that cause decline and reserve factors that compensate. *Neuron, 44*(1), 195–208. https://doi.org/10.1016/j.neuron.2004.09.006

[11] Carthery-Goulart, M. T., Privitera, A. J., & Weekes, B. S. (2023). Does language distance modulate the contribution of bilingualism to cognitive reserve in seniors? A systematic review. *American Journal of Alzheimer's Disease & Other Dementias, 38*, 1–17. https://doi.org/10.1177/15333175231167223

[12] Chan, C. K. Y., Wong, R. H. M., & Li, Y. (2022). Cognitive engagement in mobile-assisted language learning: Links to executive function components. *Journal of Computer-Assisted Language Learning, 35*(3), 311–334.

[13] Demnitz-King, H., Requier, F., Whitfield, T., Schlosser, M., Gonneaud, J., Ware, C., Barnhofer, T., Coll-Padros, N., Dautricourt, S., Delarue, M., Klimecki, O. M., Paly, L., Salmon, E., Schild, A.-K., Wirth, M., Frison, E., Lutz, A., Chételat, G., Collette, F., … Marchant, N. L. (2023). Effects of meditation training and non-native language training on cognition in older adults: A secondary

analysis of a randomized clinical trial. *JAMA Network Open, 6*(7), e2317848. https://doi.org/10.1001/jamanetworkopen.2023.17848

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

[15] Furlan, A. D., Pennick, V., Bombardier, C., & van Tulder, M. W. (2009). 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine, 34*(18), 1929–1941. https://doi.org/10.1097/BRS.0b013e3181b1c99f

[16] Gonzales, M. M., Garbarino, V. R., Pollet, E., Palavicini, J. P., Kellogg, D. L., Jr., Kraig, E., & Orr, M. E. (2022). Biological aging processes underlying cognitive decline and neurodegenerative disease. *Journal of Clinical Investigation, 132*(10), e158453. https://doi.org/10.1172/JCI158453

[17] Grossmann, J. A., Aschenbrenner, S., Teichmann, B., & Meyer, P. (2023). Foreign language learning can improve response inhibition in individuals with lower baseline cognition: Results from a randomized controlled superiority trial. *Frontiers in Aging Neuroscience, 15*, 1123185. https://doi.org/10.3389/fnagi.2023.1123185

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

[19] Hertzog, C., Kramer, A. F., Wilson, R. S., & Lindenberger, U. (2008). Enrichment effects on adult cognitive development: Can the functional capacity of older adults be preserved and enhanced? *Psychological Science in the Public Interest, 9*(1), 1–65. https://doi.org/10.1111/j.1539-6053.2009.01034.x

[20] Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2024). Cochrane handbook for systematic reviews of interventions (Version 6.5). Cochrane. https://training.cochrane.org/handbook

[21] Hou, J., Jiang, T., Fu, J., Su, B., Wu, H., Sun, R., & Zhang, T. (2020). The long-term efficacy of working memory training in healthy older adults: A systematic review and meta-analysis of 22 randomized controlled trials. *The Journals of Gerontology: Series B, 75*(8), e174–e188. https://doi.org/10.1093/geronb/gbaa077

[22] Hozo, S. P., Djulbegovic, B., & Hozo, I. (2005). Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology, 5*, 13. https://doi.org/10.1186/1471-2288-5-13

[23] Kliesch, M., Becker, R., Hervais-Adelman, A., Pfenninger, S. E., & Meyer, M. (2022). Global and localized network characteristics of the resting brain predict and adapt to foreign language learning in older adults. *Scientific Reports, 12*, 3633. https://doi.org/10.1038/s41598-022-07629-y

[24] Klimova, B., Slaninova, G., Prazak, P., Kacetl, J., & Valis, M. (2020). Enhancing cognitive performance of healthy Czech seniors through non-native language learning: A mixed-methods pilot study. *Brain Sciences, 10*(9), 573. https://doi.org/10.3390/brainsci10090573

[25] Knowles, M. S., Holton, E. F., III, & Swanson, R. A. (2015). The adult learner: The definitive classic in adult education and human resource development (8th ed.). Routledge.

[26] Lampit, A., Hallock, H., & Valenzuela, M. (2014). Computerized cognitive training in cognitively healthy older adults: A systematic review and meta-analysis of effect modifiers. *PLOS Medicine, 11*(11), e1001756. https://doi.org/10.1371/journal.pmed.1001756

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

[28] Livingston, G., Huntley, J., Liu, K. Y., Costafreda, S. G., Selbæk, G., Alladi, S., Ames, D., Banerjee, S., Burns, A., Brayne, C., Fox, N. C., Ferri, C. P., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Nakasujja, N., Rockwood, K., … Mukadam, N. (2024). Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. *The Lancet, 404*(10452), 572–628. https://doi.org/10.1016/S0140-6736(24)01296-0

[29] Meltzer, J. A., Rose, M. K., Le, A. Y., Spencer, K. A., Goldstein, L., Gubanova, A., Lai, A. C., Yossofzai, M., Armstrong, S. E. M., & Bialystok, E. (2023). Improvement in executive function for older adults through smartphone apps: A randomized clinical trial comparing language learning and brain training. *Aging, Neuropsychology, and Cognition, 30*(2), 150–171. https://doi.org/10.1080/13825585.2021.1991262

[30] Ngandu, T., Lehtisalo, J., Solomon, A., Levälahti, E., Ahtiluoto, S., Antikainen, R., Bäckman, L., Hänninen, T., Jula, A., Laatikainen, T., Lindström, J., Mangialasche, F., Paajanen, T., Pajala, S., Peltonen, M., Rauramaa, R., Stigsdotter-Neely, A., Strandberg, T., Tuomilehto, J., … Kivipelto, M. (2015). A 2-year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): A randomised controlled trial. *The Lancet, 385*(9984), 2255–2263. https://doi.org/10.1016/S0140-6736(15)60461-5

[31] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

[32] Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: Aging and neurocognitive scaffolding. *Annual Review of Psychology, 60*, 173–196. https://doi.org/10.1146/annurev.psych.59.103006.093656

[33] Parmaxi, A., & Zaphiris, P. (2017). Gamification in computer-assisted language learning: A systematic review. *Computer Assisted Language Learning, 30*(5), 448–472.

[34] Paulavicius, E., Mizzaci, C. C., Tavares, D. R. B., Rocha, A. P., Civile, V. T., Schultz, R. R., Pinto, A. L., & Trevisani, V. F. M. (2020). Bilingualism for delaying the onset of Alzheimer's disease: A systematic review and meta-analysis. *European Geriatric Medicine, 11*(4), 651–658. https://doi.org/10.1007/s41999-020-00326-x

[35] Petersen, R. C. (2011). Mild cognitive impairment. *New England Journal of Medicine, 364*(23), 2227–2234. https://doi.org/10.1056/NEJMcp0910237

[36] Pliatsikas, C. (2020). Understanding structural plasticity in the bilingual brain: The Dynamic Restructuring Model. *Bilingualism: Language and Cognition, 23*(2), 459–471. https://doi.org/10.1017/S1366728919000130

[37] Ramos, S., Fernandez Garcia, Y., Anton, E., Casaponsa, A., & Dunabeitia, J. A. (2017). Does learning a language in the elderly enhance switching ability? *Journal of Neurolinguistics, 43*, 39–48. https://doi.org/10.1016/j.jneuroling.2016.09.001

[38] Reuter-Lorenz, P. A., & Park, D. C. (2014). How does it STAC up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychology Review, 24*(3), 355–370. https://doi.org/10.1007/s11065-014-9270-9

[39] Stern, Y. (2009). Cognitive reserve. *Neuropsychologia, 47*(10), 2015–2028. https://doi.org/10.1016/j.neuropsychologia.2009.03.004

[40] Stine-Morrow, E. A. L., & Manavbasi, I. E. (2022). Beyond "use it or lose it": The impact of engagement on cognitive aging. *Annual Review of Developmental Psychology, 4*, 319–352. https://doi.org/10.1146/annurev-devpsych-121020-030017

[41] Valenzuela, M. J., & Sachdev, P. (2006). Brain reserve and dementia: A systematic review. *Psychological Medicine, 36*(4), 441–454. https://doi.org/10.1017/S0033291705006264

[42] Ware, C., Dautricourt, S., Gonneaud, J., & Chételat, G. (2021). Does second language learning promote neuroplasticity in aging? A systematic review of cognitive and neuroimaging studies. *Frontiers in Aging Neuroscience, 13*, 706672. https://doi.org/10.3389/fnagi.2021.706672

[43] Wong, P. C. M., Ou, J., Pang, C. W. Y., Zhang, L., Tse, C. S., Lam, L. C. W., & Antoniou, M. (2019). Language training leads to global cognitive improvement in older adults: A preliminary study. *Journal of Speech, Language, and Hearing Research, 62*(7), 2411–2424. https://doi.org/10.1044/2019_JSLHR-L-18-0321

[44] World Health Organization. (2026). Risk reduction of cognitive decline and dementia: WHO guidelines (2nd ed.). World Health Organization.

Funding

The author(s) disclosed this research did not receive funding.