《心理科学进展》审稿意见与作者回应

题目：老年阶段外语学习改善执行功能？一项三水平元分析
作者：张楠，黄立鹤

第一轮

审稿人 1 意见：
意见 1：

关于引言叙述非常啰嗦，尽量压缩为一小段，迅速精简的聚焦于本话题的研究现状，研究意义和以往研究的不足并提出自己的研究目标。叙述时需要注意，本研究关注的话题为老年人，所以引文及文字叙述应该更多的围绕该群体展开。

回应：

非常感谢审稿专家的细致审阅！引言已按意见凝练为一段，逻辑链调整为：老龄化危机 → 非药物干预需求 → 外语学习干预的神经科学与现实依据 → 研究现状及不足 → 研究目标与意义，由大到小，迅速聚焦。同时，全段主语与引文均严格围绕老年人展开，确保服务于老年人认知干预这一核心主题。以上修改恳请您再次审阅。

意见 2：

引言尤为需要说明本研究的现实意义，比如当下时代启动外语学习的群体有多少，趋势怎么样，是不是一个值得关注的社会现象。单纯为了改善执行功能，老年学习外语，可操作性的空间有多强。几近退休的老人们是更愿意喝茶，下棋，锻炼，还是更愿意关起门来学外语。作者需要交代。否则本文即便有统计意义也不是一个好的现实话题。

回应：

非常感谢审稿专家对引言撰写思路的耐心指导！经反复思考，我们从**趋势、需求与动机**三个层面补充现实意义。

首先，关于老年外语学习“是不是一个值得关注的社会现象”这一问题，鉴于老年外语学习渠道碎片化(老年大学、线上平台、APP 自学等)，参与规模难以估计，我们转而补充提及老年人外语学习参与热情日益增长**趋势**的相关文献(Fan & Chiang, 2025; Kacetl & Klímová, 2021)，表明该现象已值得重视(见引言标绿文字)。

其次，关于老人们“更愿意喝茶，下棋，锻炼，还是更愿意关起门来学外语”，引言中已补充外语学习相较于传统娱乐及益智活动的独特之处：为老年人提供了终身学习和开阔视野的机会，且语言学习过程本身亦可能在更广泛意义上惠及老年人的整体福祉(Brouwer et al., 2025b)更契合新一代退休老人从“生存型”到“发展型”的养老**需求**转变(见引言标红文字)。

最后，关于“单纯为了改善执行功能，老年学习外语，可操作性的空间有多强”，我们补充说明：在老年人学习外语的**动机**研究中，“健脑延年”因素排在第一位(季传峰, 2015; 张军, 2013)。这说明老年人有较强意愿通过外语学习改善认知能力，该干预手段具有一定的可操作空间(可见引言标蓝文字)。

参考文献

季传峰. (2015). 积极老龄化视角下的中老年外语学习动机研究. *当代继续教育, 33*(5), 72–76.

张军. (2013). 老年人英语学习动机研究. *语文学刊(外语教育教学)*, (11), 90–92.

Brouwer, J., van den Berg, F., Knooihuizen, R., Loerts, H., & Keijzer, M. (2025b). 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.

Fan, L., & Chiang, J. (2025). Unveiling the English proficiency of Chinese older adults through a national survey. *Educational Gerontology*, *51*(4), 387–401.

Kacetl, J., & Klímová, B. (2021). Third-age learners and approaches to language teaching. *Education Sciences*, *11*(7), 310.

意见 3：

关于 1.1 老年阶段外语学习与执行功能的关系。叙述颇为单薄，缺少必要的理论基础，仅从实证研究结果上简单提及了不同的发现。这是远远不够的。而，1.3 又显得非常突兀。建议 1.3 部分内容与 1.1 部分合并，适合放在调节效应中的放到合适的调节变量处。

此外，主效应这块作者应该有倾向性的假设两者之间是否存在关系，是直线的还是曲线的。是正的还是负的。否则后文调节效应的假设无从谈起。

回应：

非常感谢审稿专家的宝贵意见！我们认真阅读并逐条进行了修改。

首先，针对“1.1 部分叙述较为单薄，缺少必要的理论基础”这一点，我们对该部分进行了理论补充。具体而言，我们补充了执行功能的经典定义、神经生理学视角下的前额叶老化假说以及认知心理学层面的认知储备假说，系统阐述了老年外语学习与执行功能的关系。这一修改烦请您查看 **1.1 节的第一段整段**。

其次，我们将原 1.3 节中“干扰抑制效应与加工复杂性效应”及“BICA 与 BEPA 假说”两部分内容做了删减，并将其自然融入 1.1 节，置于理论基础(即 1.1 节第一段)之后，作为作用机制(外语学习输入端与认知表现输出端)加以呈现，详见 1.1 节第二段标红处。由于这两对假说的检验分别与“语言距离”与“执行功能子成分”两个调节变量的检验相关，假说的内容也已移至各自对应的调节变量部分，详见 1.2.1 节的“语言距离”和 1.2.3 节的“执行功能子成分”的标蓝处。

最后，我们在 1.1 节结尾明确了主效应倾向性假设：老年阶段外语学习对执行功能存在显著正向主效应，且该效应并非简单的线性关系，为后续调节效应假设提供了支撑。详细修改请您查看 1.1 节最后一段的标蓝处。上述修改显著提升了本研究的理论深度，再次感谢审稿专家的专业意见！

意见 4：

“其次，在研究维度上，不仅细化变量设置(如增设“认知状态”变量)，更突破既有元分析侧重于干预与受试特征的局限，通过增设结果特征维度，系统分析执行功能各核心成分的改善情况，在研究广度与深度上均有所拓展”。作者需要解释清楚所谓的认知状态和结果特征。

回应：

非常感谢审稿专家的宝贵意见！由于原引言部分已作凝练，本条意见中的相关表述已从引言部分移至 **1.2 节整节的最后一段**(见该段标绿字体)。为更清晰地回应您的关切，我们在修改稿中对这两个新增维度进行了详细阐述，并在相应的 1.2.2 节“认知状态”这一部分和 1.2.3 节“执行功能核心子成分”部分对“认知状态”和“结果特征”分别做出了清晰界定(详见这两部分第一段的标红字体)。具体说明如下：

首先，关于“认知状态”变量：该变量指老年参与者的基线认知功能水平，我们将其细化为两个水平：(1)认知正常老年人；(2)认知障碍。需要注意的是，本研究将“认知障碍”人群定义为客观认知功能受损但尚未达到重度痴呆的群体。这是因为该群体虽存在认知障碍，但仍具备外语学习的可行性。这一变量的增设突破了以往元分析中对老年受试认知状态特征的

忽略，能够更精细地考察老年外语学习的效果是否因基线认知状态不同而存在异质性。

其次，增设“结果特征”维度，是因为既有元分析多聚焦于干预特征和受试特征，而较少对结果变量本身进行系统拆解。本研究参照 Diamond(2013)将执行功能细分为其三个核心成分——抑制控制、工作记忆和认知灵活性，分别计算并比较各子成分的效应量，便于厘清既有研究结论不一致的可能来源，同时也使 BICA 与 BEPA 两大假说支持情况的考察可操作化(具体操作方法见 2.3.1 标绿字体)。

参考文献

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(1), 135–168.

意见 5：

所有的调节变量的论述都缺少理论支撑，并且没有根据理论观点或者已有研究结果提出自己的倾向性假设。

回应：

非常感谢审稿专家提出的宝贵意见！原稿未能为调节变量的提出做好理论支撑，也未清晰提出基于理论或实证的倾向性假设。为此，我们对所有调节变量补充了明确的理论支撑(烦请您查看每个调节变量部分的标蓝字体)，并严格依据这些理论观点和已有实证结果，提出了具有明确倾向性的研究假设(烦请您查看每个调节变量部分的标绿字体)。

各调节变量的具体修改思路如下：

语言距离：原稿仅描述既有研究分歧，缺少理论根源。现新增 Antoniou 和 Wright (2017)提出的“干扰抑制效应”与“加工复杂性效应”两大互竞理论假说作为支撑，并结合 Shi 等人(2025)的动态视角，提出倾向性假设：语言距离对执行功能增益存在显著调节作用，且该效应具有阶段依赖性——远距离语言在短期干预初期更具优势。

学习时长与学习强度：原稿仅呈现实证争议。现新增认知训练领域的剂量-反应模型作为理论支撑，并依据 Antoniou 等人(2013)对高强度训练的强调，提出倾向性假设：学习总时长与学习强度对执行功能增益均存在非线性调节作用。具体而言，随着总学习时长的累积或学习强度的提升，干预阶段前期的认知获益增长较为迅速，后期则表现出明显的边际效益递减(即收益曲线趋于平缓或达到平台期)。

对照类型：原稿仅描述研究设计差异。现以“语言特异性 vs 非特异性效应”争议为理论支撑，明确外语学习的语言竞争、一般认知激活、社交参与三重属性，提出倾向性假设：剥离语言学习的非特异性贡献后，语言竞争特异性增益仍显著，但“外语学习-无干预”组效应量(综合效应量)最大。

年龄：原稿仅提及年龄分层。现引入 Lövdén 等人(2010)的成人认知可塑性理论框架，强调随年龄增长可塑性容量下降，提出倾向性假设：年龄对老年外语学习执行功能增益存在显著负向调节作用，年龄越大，增益越小。

认知状态：原稿仅引用 Kliesch 等人(2022)的实证研究。现以 Lövdén 等人(2012)提及的补偿观(compensation account)为理论支撑(预测初始认知能力低的个体获益更多)，结合 Kliesch 等人(2022)的证据，提出倾向性假设：低基线认知状态(认知障碍)的老年人将比认知正常老年人获得更大的执行功能改善。

执行功能核心子成分：原稿仅列举实证证据。现溯源至 BICA 与 BEPA 两大假说之争作为理论基础，明确指出不同子成分敏感度差异的理论根源，并提出倾向性假设：三个核心子成分的增益无显著差异，支持 BEPA 假说，且该结果与对“语言距离”调节效应的倾向性假设逻辑一致：远距离语言的学习在短期(初期)干预中更具优势，因加工复杂性更高，更容易引发执行功能的整体优化。

我们对本次修改非常重视，如仍有任何不足，恳请审稿专家继续指正，我们将进一步完

善。再次衷心感谢您的意见！

意见 6：

PRISMA 指南有 2020 新版的，作者可参考。

回应：

非常感谢审稿专家的宝贵意见！我们已参照 PRISMA 2020 新版指南(Page et al., 2021)对元分析报告进行了修订。本次修改重点补充了 2020 版更新的报告要求：在文献检索部分，完整列出了所有数据库及其他来源，明确了最终检索日期(详见 2.1 节)；在附录中提供了各数据库的完整检索策略(详见附录-附表 1)；文献筛选流程图已按 PRISMA 2020 推荐模板重新绘制，并标注了各阶段纳入与排除的文献数量(详见 3.1 节的图 1)。上述修改使论文的可复现性得到提升，再次感谢审稿专家的指导！

参考文献

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372,* Article n71. https://doi.org/10.1136/bmj.n71

意见 7：

“Web of Science，PubMed，ProQuest 等外文数据库中，”此处不适合用等，有几个数据库就如实报告几个。

回应：

非常感谢审稿专家的专业指正！我们已对文献检索部分的表述进行了修改。原句“Web of Science，PubMed，ProQuest 等外文数据库中”和“在知网、维普、万方等中文数据库中”两处均已重新表述，详见 2.1 节标红字体。修改后文献检索部分的表述更加规范，感谢您的指导！

意见 8：

作者是在哪进行的关键词检索，摘要、全文还是标题？

回应：

非常感谢审稿专家的提问！在修改稿中，我们补充了详细的文献检索策略。检索在标题、摘要及关键词字段中进行。限于数据库自身的技术架构、字段定义和检索语法差异，各数据库的检索条件无法完全统一，具体为：Web of Science 使用主题字段(覆盖标题、摘要及关键词)，PubMed 和 ProQuest 限定标题和摘要字段，知网和维普采用篇名、关键词和摘要字段，万方数据库限摘要字段。

以上修改我们已在 2.1 中详细写明(见标绿字体)，并在附录的附表 1 中提供了所有数据库的完整检索策略，恳请审稿专家再次审阅！

意见 9：

作者的检索词不具有互斥性，而是存在包含与被包含的关系，也就是说，如果作者搜了“学习” OR “习得”那就没必要再搜“语言学习” OR “语言习得”。以此类推。所以建议作者重新合理化检索词，否则对于读者和未来的元分析操作者而言是一种严重的误导。

回应：

非常感谢审稿专家指出检索词存在的问题！我们重新审视了文献检索部分的检索词。原检索词确实存在包含与被包含的层级关系，造成了不必要的冗余。为此，我们对检索词进行了优化：删除了冗余的包含性术语，优化了布尔逻辑的组合方式，将检索词划分为三个相互独立的概念组(见下图中的 S1, S2, S3)，再进行组合检索。同时，我们保留了必要的同义词

和截词符以确保检索敏感性，具体调整如图所示(以知网为例)。

知网	S1	篇关摘:(老年 + 老人 + 老者 + 长者 + 晚年 + 老化)
	S2	篇关摘:(((二语 + 外语) * (学习 + 习得)) + 语言干预 + 语言培训 + 语言训练)
	S3	篇关摘:(认知 + 执行功能 + 工作记忆 + 抑制控制 + 认知灵活性 + 转换 + 切换)
	检索策略	**S1 AND S2 AND S3**

图 1 检索策略(以知网为例)

我们已在附录的**附表 1** 中详细列出了所有数据库的完整检索策略，便于读者和后续研究者复现，再次感谢审稿专家的细致审阅！

意见 10：

作者的检索时间为 2025 年 11 月 30 日。最近一次更新是什么时候？

回应：

非常感谢审稿专家的提问！原稿文献检索的最后更新时间即为 2025 年 11 月 30 日。由于该时间距今已有半年，确有必要进行更新。因此，本研究补充检索了 2025 年 12 月 1 日至 2026 年 4 月 30 日期间的相关文献(2.1 节表述已更新，见橙色字体)。本次检索共发现 3 篇符合纳入标准的文献。第 1 篇(Bubbico et al., 2026)的原始数据已发表于作者 2019 年的文献中，已被本研究纳入，不再重复纳入。第 2 篇(李想 等, 2025)与第 3 篇(Che & Huang, 2026)符合要求，作为新文献纳入。

此外，针对下方意见 15 和意见 17 的回应涉及数据分析方法的调整，我们将分析工具由 CMA 更换为 R。R 能够兼容更多数据类型，因此先前因数据类型不符合 CMA 要求而被排除的三项研究(Meltzer et al., 2023; Schultz et al., 2024; van der Ploeg et al., 2023)得以重新纳入。综合以上调整，更新后共补充纳入 5 篇文献(含上述重新纳入的 3 篇与新检索到的 2 篇)，详细编码表见下方对意见 17 的回应。文稿其他部分中，凡涉及上述修改之处均已同步更新。

参考文献

李想, 冯燕, 张语静, 储艺涵. (2025). 基于外语学习的老年人认知训练研究. *心理学进展*, *15*(12), 36–44. https://doi.org/10.12677/ap.2025.1512625

Bubbico, G., Tomaiuolo, F., Sestieri, C., Akhlaghipour, G., Granzotto, A., Ferretti, A., Perrucci, M. G., Sensi, S. L., & Delli Pizzi, S. (2026). Foreign language learning in older adults modifies resting-state functional connectivity between the subcortical structures and the cortex. *Aging Medicine, 9*(2), 99–110. https://doi.org/10.1002/agm2.70073

Che, Y., & Huang, L. (2026). Foreign language learning for active aging: Cognitive and psychosocial benefits of Chinese older adults learning English. *Educational Gerontology*. Advance online publication. https://doi.org/10.1080/03601277.2026.2647775

Meltzer, J. A., Kates Rose, M., 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

Schultz, D. H., Gansemer, A., Allgood, K., Gentz, M., Secilmis, L., Deldar, Z., Savage, C. R., & Ghazi Saidi, L. (2024). Second language learning in older adults modulates Stroop task performance and brain activation. *Frontiers in Aging Neuroscience*, *16*, Article 1398015. https://doi.org/10.3389/fnagi.2024.1398015

van der Ploeg, M., Lowie, W., & Keijzer, M. (2023). The effects of language teaching pedagogy on cognitive functioning in older adults. *Behavioral Sciences*, *13*(3), Article 199. https://doi.org/10.3390/bs13030199

意见 11：

作者的检索时间为 2025 年 11 月 30 日。审稿人发现作者的预注册实践并不是再此之前，而是最近几天的事情。这怎么能算是预注册呢？所以自检报告中作者不应该将本研究视为预注册过。希望作者能够带着严谨的态度打磨文章，真正的好文章不用因此而心虚，审稿人也会不拘小节。

回应：

非常感谢审稿专家的严格把关！我们已认真核查时间线：原稿文献检索完成于 2025 年 11 月 30 日，而预注册登记确实晚于检索日期，显然不符合预注册的定义(应在研究开展前完成注册)及其防止重复研究的目的。我们诚恳接受审稿专家的批评并为此深表歉意。据此，我们已修改“元分析自检报告”问题 5 的回答，明确写明“未进行预注册”。

我们会以此为鉴，在今后的研究中严格遵守预注册等最佳实践。衷心感谢审稿专家的认真审阅、善意提醒与宽厚包容，我们将以更加严谨的态度继续打磨论文！

意见 12：

“(5)支持的假说”这也要编码吗？

回应：

感谢审稿专家的提问，关于假说的编码我们作出如下解释。

就本元分析的研究问题而言，涉及“干扰抑制效应与加工复杂性效应”和“BICA 假说与 BEPA 假说”两对假说的支持性检验，检验逻辑如下：

干扰抑制效应与加工复杂性效应(外语学习输入端两大互竞假说)**需要编码**。编码逻辑为：若某文献的实验设计中目标语与母语同属一语系(语言距离较近)，则该文献的效应值被编码为“支持干扰抑制效应假说”；若目标语与母语不属一语系(语言距离较远)，则编码为“支持加工复杂性效应假说”。后续将分别对支持两个假说的效应值进行合并，得出各自的代表性效应值总值，从而量化检验两大假说的支持力度。

BICA 假说与 BEPA 假说(认知表现输出端两大互竞假说)**不需要编码**。这两大假说的验证逻辑依赖于“执行功能核心子成分”的亚组分析：若抑制控制的效应量显著优于工作记忆和认知灵活性，则支持 BICA 假说；若三个子成分之间的亚组差异不显著，则支持 BEPA 假说(即执行加工的整体优化)。因此，我们将通过执行功能核心子成分这一调节变量的亚组分析直接检验 BICA 与 BEPA 假说的相对支持情况，而无需为其单独设置编码类别。

以上检验逻辑已在 2.3.1 最后一段详细阐述(见 2.3.1 标绿字体)，恳请审稿专家再次审阅！

意见 13：

“文献质量评估参考曹贤才等人(2025)针对老年研究修订的标准：”作者的引用不严谨。文献质量评价标准并不是曹贤才制定的，而是曹贤才引文中的张亚利制定的。该文并未说明标准的合理性，比如为何博士论文为 2 分，硕士论文为 1 分？建议作者采用国际上更为通行的统一客观标准衡量文献质量，而不是私自做主，随意改编。

回应：

非常感谢审稿专家的细致审阅！经仔细核查，原稿中“文献质量评估参考曹贤才等人(2025)针对老年研究修订的标准”确实表述不严谨。修改稿中，我们采用国际通行的美国国立卫生研究院(National Institutes of Health, NIH)推荐的文献质量评估工具，对元分析纳入的所有文献重新进行了严格评估。

本研究纳入的文献涉及三种研究设计类型：一是随机对照试验(RCT)；二是非随机对照试验(nRCT，有对照组但未采用随机分配)；三是单组前后测设计(before-after/pre-post 设计)。其中，前两类研究(RCT 和 nRCT)采用 *Quality Assessment of Controlled Intervention Studies* 工

具(National Institutes of Health, 2021a)进行评估，后一类研究采用 *Quality Assessment Tool for Before-After (Pre-Post) Studies With No Control Group* 工具(National Institutes of Health, 2021b)进行评估。NIH 质量评估工具将文献质量分为三个等级：Good(良好)、Fair(一般)和 Poor(差)。为确保评估的客观性和可重复性，所有文献的质量评价均由两名编码者独立编码完成。上述修改烦请您查看 2.2 节的标蓝字体。

编码者间一致性检验结果显示，RCT 和 nRCT 类文献的 Cohen's Kappa 值为 0.758，前后测类文献的 Cohen's Kappa 值为 0.799，均显示出良好的编码者间一致性。所有最终纳入的研究均未被评估为"Poor(差)"，因此未有研究因文献质量过低而被剔除。文献质量评估结果的报告烦请您查看 3.1 的标蓝字体。

此外，我们已将元分析纳入的所有文献的研究设计类型和具体质量评分结果补充至修改稿**附表 2**，元分析自检报告题目 3 的回答也进行了相应修改，恳请审稿专家再次审阅！

需要说明的是，在质量评估过程中，我们发现原纳入的 Liu(2023)为回顾性观察研究，并非严格意义上的干预研究。因此，本次修订中已将该文献剔除，数据分析也未包含此篇。

参考文献

Liu, H. (2023). Attention: The cognitive effects of learning to read in Arabic by Chinese learners at an old age. *Sage Open*, *13*(3), 1–13. https://doi.org/10.1177/21582440231193330

National Institutes of Health. (2021a). *Quality assessment of controlled intervention studies*. National Heart, Lung, and Blood Institute. https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

National Institutes of Health. (2021b). *Quality assessment tool for before-after (pre-post) studies with no control group*. National Heart, Lung, and Blood Institute. https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

意见 14：

"照张楠和李旭奎(2026)的方法进"这些都属于不当引用，是二手引用。

回应：

非常感谢审稿专家指出这一问题。经核查，原稿中"参照张楠和李旭奎(2026)的方法进行效应量的计算与合并"的表述确实属于二手引用，已予以删除。我们重新梳理了效应量计算的具体操作，并采用以下更合适的文献进行说明。由于本研究纳入的数据共有三种类型(均值、*t* 值、Cohen's *d* 值)，因此，修订后的表述清晰写明"多数研究通过均值、标准差、样本量计算效应量的大小，少量研究将 *t* 值、Cohen's *d* 值转换为 *g* 值(Harrer et al., 2021)。"修订之处辛苦您查看 2.3.1 的标蓝字体。

参考文献

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003107347

意见 15：

研究共纳入有效文献 16 篇(含 57 个独立效应值)，审稿人查看了编码表对 57 个所谓的独立效应值深感担忧。所谓独立性是针对被试群的，而不是变量的或研究的。统一研究过系列子研究都是一堆被试，那只能是一个独立效应值，如果纳入多个结果那必然是不独立的。这种情况需要用多水平元分析处理。所以本文的结果都是错误的。

回应：

非常感谢审稿专家指出本研究在数据分析上的问题。本研究的确存在从同一研究中提取多个效应量的情况，违背了传统元分析所需效应量之间相互独立的前提假设。然而，我们之所以仍采用传统两水平元分析，是基于以下统计检验结果：首先，我们检验了水平 2 的异质

性，发现研究内部(水平 2)效应量的变异不显著(如下图，LRT 的 p 值为 0.85)。其次，我们比较了两水平模型与三水平模型的拟合优度，结果如下图所示：Full 代表储存于变量 overallgp 的三水平模型，Reduced 代表存储于变量 modelnolev2 的两水平模型。Reduced 模型的 AIC 和 BIC 均小于 Full 模型，表明两水平的模型拟合优于三水平模型。

```
> # 水平2异质性检验
> modelnolev2 <- rma.mv(gi,
+                       vi,
+                       random = list(~ 1 | esID, ~ 1 | studyID),
+                       sigma2 = c(0, NA),
+                       test = "t",
+                       data = X3水平元分析数据)
>
> anova(overallgp, modelnolev2)

          df     AIC     BIC    AICc   logLik    LRT   pval       QE
Full       3 32.3298 38.4058 32.7913 -13.1649                 115.9391
Reduced    2 30.3640 34.4147 30.5905 -13.1820 0.0343 0.8531 115.9391
```

图 2 模型拟合优度

在审稿专家指出数据分析上的问题以后，我们进一步查阅了相关文献，发现有学者提出，尽管统计结果倾向于更简单的两水平模型，保留三水平模型仍是有必要的，因为它更真实地反映了数据实际产生的过程(data-generating process) (Harrer et al., 2021：298)。因此，我们采用三水平元分析的方法重新进行了数据分析。相关的发表偏倚检验、敏感性分析、异质性检验等都进行了更新。由于研究内部(水平 2)的变异不显著，因此本次三水平元分析的结果与原本两水平元分析的结果差异不大，恳请审稿专家继续审阅数据分析以及后续的结果与讨论部分，感谢您的耐心！

此外，由于我们采用 R 进行新一轮的数据分析，可纳入的数据类型不受原本 CMA 软件的限制，因而包容度更强。所以先前因为数据类型不符而被剔除的 3 篇文献(Meltzer et al., 2023; Schultz et al., 2024; van der Ploeg et al., 2023)被重新纳入此次分析。结合意见 10，我们补充检索了 2025 年 12 月 1 日至 2026 年 4 月 30 日期间发表的文献，纳入 2 篇符合标准的新文献(Che & Huang, 2026；李想 等, 2025)。综上所述，本轮修改共新纳入 5 篇文献，编码信息见下方表 1。

表 1　新纳入文献编码表

作者与发表年份	样本量	干预特征			受试特征			结果特征	支持的假说	效应量	文献质量
		语言距离	学习时长	学习强度	对照类型	年龄	认知状态	执行功能核心子成分			
Che & Huang, 2026	24	远	38.00	/	/	66.90	认知健康	工作记忆	加工复杂性效应	0.79	Fair
	24	远	38.00	/	/	66.90	认知健康	/(执行功能整体)	加工复杂性效应	0.38	
Meltzer et al., 2023	52	近	40.00	12.50	外语学习-无干预	69.80	认知健康	工作记忆	干扰抑制效应	0.39	Fair
	52	近	40.00	12.50	外语学习-无干预	69.80	认知健康	认知灵活性	干扰抑制效应	0.49	
Schultz et al., 2024	41	/	81.40	27.7	/	66.63	认知健康	抑制控制	/	0.49	Fair
van der Ploeg et al., 2023	16	近	84.00	35.00	/	71.9	认知健康	工作记忆	干扰抑制效应	0.32	Fair
李想 等, 2025	33	远	8	4	外语学习-主动对照	63.74	/	工作记忆	加工复杂性效应	0.03	Fair
	33	远	8	4	外语学习-主动对照	63.74	/	工作记忆	加工复杂性效应	0.18	
	28	远	8	4	外语学习-被动对照	63.74	/	/	加工复杂性效应	1.29	
	28	远	8	4	外语学习-被动对照	63.74	/	/	加工复杂性效应	1.19	

参考文献

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003107347

意见 16：

图 1 文献纳入和排除过程 请采用 PRISMA 官网规范的报告流程图。

回应：

非常感谢审稿专家的耐心指导！我们已按照 PRISMA 官网规范的流程图格式，重新绘制了文献纳入与排除流程图，如图 3 所示。该图已在修改稿中相应位置(3.1 节的图 1)更新，恳请您再次审阅！

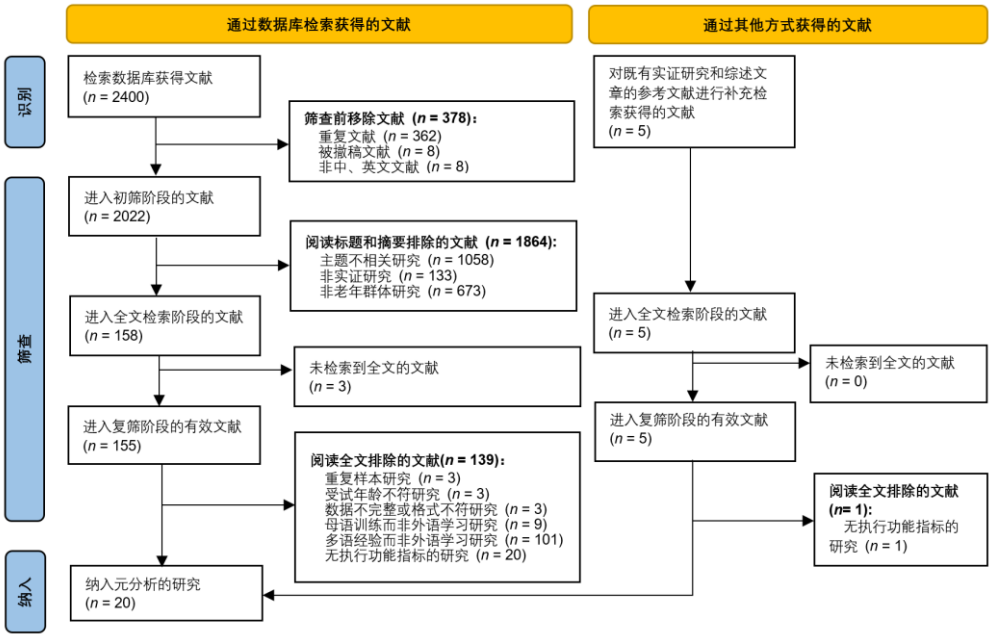


图 3 文献纳入与排除流程图

意见 17：

由于数据分析上的错误，结果和讨论部分目前是无意义的。

回应：

非常感谢审稿专家指出本研究数据分析中存在的问题。我们已在意见 15 的回应中对相关问题进行了详细说明，并对数据分析方法进行了重新审视与修正。目前已完成重新分析，统计分析方法(**第 2.3 节**)已据此修订，结果与讨论部分也已基于新的分析结果进行了更新。敬请审稿专家拨冗重新审阅修订后的第 2.3 节“统计分析”、“结果”及“讨论”部分，感谢您的耐心！

……

审稿人 2 意见：

该元分析探讨老年阶段启动的外语学习对老年人执行功能的影响，选题有意义。目前还存在如下一些问题。

意见 1：

老年阶段学习外语的实证研究很难做，能够凑齐 16 项不容易。但是，干预效应的金标准是 RCT，建议作者介绍 16 项研究有多少是 RCT。

回应：

非常感谢审稿专家对本研究文献检索工作的理解与肯定，也感谢您提出的关于 RCT 数量这一重要问题！需要说明的是，本轮修改过程中，因多种原因纳入文献的总量发生了变化，现向您汇报如下：(1)本轮补充检索了 2025 年 12 月 1 日至 2026 年 4 月 30 日期间的相关文献，发现 3 篇符合纳入标准的新文献。(2)结合另一位审稿专家的意见，本轮修改涉及数据分析方法的调整，我们将分析工具由 CMA 更换为 R。R 能够兼容更多数据类型，因此先前因数据类型不符合 CMA 要求而被排除的 3 项研究(Meltzer et al., 2023; Schultz et al., 2024; van der Ploeg et al., 2023)得以纳入。(3)本轮修改涉及文献质量的重新评估。在评估过程中，我们发现原纳入的 Liu(2023)为回顾性观察研究，并非严格意义上的干预研究，故将该文献

剔除。综上所述，本研究最终纳入文献 20 篇。

经逐一核查，在这 20 篇研究中，有 11 篇(共 37 个效应量)为 RCT。以这 11 项 RCT 为元分析样本重新计算效应值，结果显示：晚年外语学习对执行功能的总体效应量为 0.29 (95% CI [0.12, 0.46], $p < 0.01$)，表明外语学习对执行功能仍具有小但显著的增益效果。我们已在修订稿编码表(附录-附表 2)中新增“研究设计”一列，逐一标注每项研究的设计类型为随机对照(RCT)、非随机对照(nRCT)或前后测设计(pre-post) (纳入文献仅涉及这三种类型)。讨论部分也已补充了上述分析结果，恳请您查看 4.1 节第 1 段的标蓝字体。再次感谢审稿专家的宝贵意见！

参考文献

李想, 冯燕, 张语静, 储艺涵. (2025). 基于外语学习的老年人认知训练研究. *心理学进展*, *15*(12), 36–44. https://doi.org/10.12677/ap.2025.1512625

Bubbico, G., Tomaiuolo, F., Sestieri, C., Akhlaghipour, G., Granzotto, A., Ferretti, A., Perrucci, M. G., Sensi, S. L., & Delli Pizzi, S. (2026). Foreign language learning in older adults modifies resting-state functional connectivity between the subcortical structures and the cortex. *Aging Medicine, 9*(2), 99–110. https://doi.org/10.1002/agm2.70073

Che, Y., & Huang, L. (2026). Foreign language learning for active aging: Cognitive and psychosocial benefits of Chinese older adults learning English. *Educational Gerontology*. Advance online publication. https://doi.org/10.1080/03601277.2026.2647775

Liu, H. (2023). Attention: The cognitive effects of learning to read in Arabic by Chinese learners at an old age. *Sage Open*, *13*(3), 1–13. https://doi.org/10.1177/21582440231193330

Meltzer, J. A., Kates Rose, M., 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

Schultz, D. H., Gansemer, A., Allgood, K., Gentz, M., Secilmis, L., Deldar, Z., Savage, C. R., & Ghazi Saidi, L. (2024). Second language learning in older adults modulates Stroop task performance and brain activation. *Frontiers in Aging Neuroscience*, *16*, Article 1398015. https://doi.org/10.3389/fnagi.2024.1398015

van der Ploeg, M., Lowie, W., & Keijzer, M. (2023). The effects of language teaching pedagogy on cognitive functioning in older adults. *Behavioral Sciences*, *13*(3), Article 199. https://doi.org/10.3390/bs13030199

意见 2：

关于对照类型，主动对照、被动对照、无干预是领域内的标准提法吗？建议作者增加三种提法的准确描述。

回应：

非常感谢审稿专家的提问与指导！关于对照类型的命名，“主动对照(active control)、被动对照(passive control)、无干预(no-training / no-intervention)”三种提法在近年来老年认知干预领域(尤其是探讨外语学习对老年认知影响的干预研究中)已逐渐成为通用的对照组分类方式。例如，Kliesch 等人(2022: 655)指出：“*Other studies used passive control conditions (PASV) for comparison…; yet others used both a passive and an active (ACTV) control group*”。该文表 2 中亦反复提及“*no training*”条件。本轮修订中，我们参照既有文献，对这三个对照组进行了更为明确的描述：

主动对照 (active control)：与外语学习组共同具备一般认知激活和社交参与属性，二者对照旨在分离这两个“非外语学习特异性”因素，突出“语言竞争”这一特异性因素的效应量。该组通常安排受试参加与外语学习组时间和强度相同、但不涉及新语言学习的其他认知活动，例

如完成电脑版填字游戏与逻辑谜题(Wong et al., 2019a)或参加乐器学习课程(Brouwer et al., 2025b)。

被动对照 (passive control)：与外语学习组共同具备社交参与属性，二者对照旨在分离这一属性(因社交可能带来心理情绪获益)。该组受试需维持与外语学习组和主动对照组程度相当的社会参与，但不施加学习负荷，如 Wong 等人(2019b)的音乐欣赏、Kliesch 等人(2022)的电影观看、Brouwer 等人(2025b)的听讲座等方式。

无干预 (no-training / no-intervention)：旨在提供老龄化认知发展的自然基线。该组受试在实验期间不接受任何人为干预，完全维持其原有日常生活。

我们已将上述描述写入修订稿 **1.2.1 小节的"(3)对照类型"的第 2 段**，恳请审稿专家查看该段落的标红字体。再次感谢您的宝贵意见！

参考文献

Brouwer, J., van den Berg, F., Knooihuizen, R., Loerts, H., & Keijzer, M. (2025b). 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.

Kliesch, M., Pfenninger, S. E., Wieling, M., Stark, A. E., & Meyer, M. (2022). Cognitive benefits of learning additional languages in old adulthood? Insights from an intensive longitudinal intervention study. *Applied Linguistics, 43*(4), 653–676.

Wong, P. C. M., Ou, J., Pang, C. W. Y., Zhang, L., Tse, C. S., Lam, L. C. W., & Antoniou, M. (2019a). Foreign language learning as potential treatment for mild cognitive impairment. *Hong Kong Medical Journal, 25*(5 Suppl. 7), S41-S43.

Wong, P. C. M., Ou, J., Pang, C. W. Y., Zhang, L., Tse, C. S., Lam, L. C. W., & Antoniou, M. (2019b). Language training leads to global cognitive improvement in older adults: A preliminary study. *Journal of Speech, Language and Hearing Research, 62*(7), 2411–2424.

意见 3：

表 2 的分类变量分析表明，"外语学习-主动对照"的研究数最多，但干预效应不显著，这与本研究目前形成的研究结论不一致，如何解释？此外，如果将"外语学习-主动对照"的 RCT 研究单独拿出来分析，是否干预的效应量更低。

回应：

非常感谢审稿专家的细致审阅，感谢您提出对"外语学习-主动对照"RCT 研究单独检验的建议！本轮修改中，由于我们更新了纳入文献与数据分析方法，表 2 的数据也已相应更新。更新后的结果显示，"外语学习-主动对照"类仍为 16 项，而"外语学习-被动对照"类增至 17 项，成为数量最多的亚组。

本轮修改中，您提出的"'外语学习-主动对照'组的干预效应不显著(g = 0.11，95% CI [−0.02, 0.23], p = 0.104)"这一现象仍存在。这一结果与主分析中外语学习整体呈现显著正向效应的结论不一致，但这恰恰体现了不同对照组类型对效应估计的不同影响。我们认为，主动对照组提供了更为严格的比较基准。该组被试接受与"外语学习组"在属性上高度匹配(均具备一般认知激活与社交参与)的其他认知活动(如填字游戏、策略游戏等)，两组间的差异效应反映的是剥离上述因素后、仅由外语学习特异性所带来的认知增益。在此严谨的对照条件下，效应量通常会缩小甚至不再显著。这并不意味着外语学习无效，而是说明其部分益处可能源于一般认知刺激或社会参与，而非主要来自外语学习本身的特异性机制(如双语激活与语言竞争)。该解释已经加入修改稿的讨论部分，恳请您查看 4.2.3 小节的标蓝部分。

此外，根据您的建议，我们对"外语学习-主动对照"中的 RCT 研究进行了单独分析。结果显示，效应量为 0.11(95% CI [−0.03, 0.24], p = 0.119)，与包含所有研究设计类型(RCT, nRCT,

pre-post)的“外语学习-主动对照”结果基本一致(g = 0.11, p = 0.104)。这一发现有两层含义：首先，即使在研究设计更为严谨的 RCT 中，外语学习相对于主动对照的优势依然不显著，表明该结果对研究质量具有一定的稳健性；其次，这也进一步支持了我们的解释，即主动对照组因能更好匹配外语学习组的非语言学习特异性因素(一般认知刺激、社会参与等)，其效应估计更为保守。即使限制在高质量 RCT 研究中，语言特异性带来的额外认知增益仍较为有限。上述补充分析及结果解释也已纳入修改稿讨论部分，烦请您查看 4.2.3 小节的标红字体。再次感谢您对本研究的耐心指导！

意见 4：

本研究发现认知障碍老人外语学习的增益效果优于健康群体，建议作者明确“认知障碍”的标准，是 SCD、MCI 还是 AD 老年人。

回应：

非常感谢审稿专家指出这一关键问题！明确“认知障碍”的具体类型对研究结论的解释至关重要。本研究将“认知障碍”人群定义为**客观**认知功能受损但尚未达到重度痴呆的群体。这是因为该群体虽存在认知障碍，但仍具备外语学习的可行性。具体而言，涉及认知障碍群体的三项研究中，Wong 等人(2019a)和陈思(2024)的受试为 MCI 患者，孙祝平等人(2024)的受试为轻度 AD 患者。相关界定已补充至修改稿 1.2.2 小节“(2)认知状态”部分(标红处)，恳请您再次审阅，感谢您的耐心！

参考文献

陈思. (2024). *基于双语认知效应的养老机构老年轻度认知障碍患者二语学习认知干预方案的构建与应用效果研究*[硕士学位论文]. 苏州大学.

孙祝平, 吴宇洁, 张卉, 徐可, 赵科. (2024). 语言训练对轻度阿尔茨海默病总体认知功能和执行功能的作用效果分析. *精神医学杂志, 37*(1), 53–58.

Wong, P. C. M., Ou, J., Pang, C. W. Y., Zhang, L., Tse, C. S., Lam, L. C. W., & Antoniou, M. (2019a). Foreign language learning as potential treatment for mild cognitive impairment. *Hong Kong Medical Journal, 25*(5 Suppl. 7), S41-S43.

第二轮

审稿人 1 意见：

意见 1：

语言上太啰嗦，全篇再精简凝练。

回应：

非常感谢审稿专家的耐心审阅！我们已对全文进行了精简凝练，主要从以下三个方面着手。**第一，压缩引言与讨论中的冗余论述。**删除了重复出现的理论背景交代、不必要的过渡句和套话，将多处长段论述合并为更为紧凑的表达。**第二，精简方法学部分的表述。**将文献检索策略、效应值计算等内容进行适度压缩，确保信息完整的同时无赘述。**第三，删除重复性内容。**对于在不同章节中反复出现的概念（如 BICA/BEPA 假说），仅保留一处详细交代，其余均以简要提及或交叉引用替代。以上修改较为琐碎，无法全部以颜色标识改动之处，还请您海涵，恳请您拨冗再次审阅！

意见 2：

3.4 和 3.7 放到调节效应部分叙述；3.5 放到主效应部分叙述。

回应：

非常感谢审稿专家的耐心指导！我们已对章节结构进行了调整：将原 3.4 节（假说的量化）与原 3.7 节（多重回归分析）移至调节效应部分，使分析逻辑更加连贯（见修改稿 3.4 节的标红字体）；同时将原 3.5 节（敏感性分析）移至主效应部分（见修改稿 3.3 节的标红字体），以增强主效应论证的稳健性。此外，将原 3.6 节（发表偏倚检验）移至主效应分析前（见修改稿 3.2 节标红字体），确保主效应分析建立在无显著发表偏倚的基础上，恳请您再次审阅！

意见 3：

4.1 第三段应放到研究意义部分阐述。

回应：

非常感谢审稿专家的细致审阅与指导！结合您提出的意见 5，我们已经将原 4.1 节的第三段移至现 4.3 节的第二段，作为本研究的“实践意义”部分，并在该段补充了“理论意义”。相关修改请您审阅现 4.3 节第二段标蓝字体，感谢您的耐心！

意见 4：

调节效应的讨论建议借鉴引言的框架，分为被试特征、干预特征等，增强整合性。

回应：

非常感谢审稿专家的细心审阅！本次修改已借鉴引言中的分类框架，将 4.2 节调节变量的讨论部分按干预特征（语言距离、学习时长、学习强度、对照类型）、受试特征（年龄、认知状态）及结果特征（执行功能核心子成分）重新组织了架构，使讨论更具整合性，恳请您再次审阅！

意见 5：

4.3 建议补充本研究的意义。

回应：

非常感谢审稿专家的建议！结合上述意见 3，我们已将原 4.1 节第三段移至现 **4.3 节第二段**作为“实践意义”，并同步补充了理论意义（见标蓝字体），形成了较为完整的研究意义阐述，恳请您再次审阅！

意见 6：

文末引文去掉超链接；很多引文的期号是错误的。

回应：

非常感谢审稿专家的细心审阅！我们已删除了文末参考文献中不必要的超链接（提前上线有 doi 的预出版论文、只有论文编号而无页码的电子刊论文需要保留超链接），并逐条核对了每篇文献的卷期号、页码等信息，确保准确无误（期号勘误之处皆已标蓝）。同时，我们已将全部参考文献格式按照《心理科学进展》文献著录要求进行了统一修改和规范。恳请您再次审阅！

意见 7：

文中括号内的引文并不是越多越好，足以论证作者的文字叙述即可，可以适当精简，能引 1 条的绝不引 2 条，以此类推。

回应：

非常感谢审稿专家的悉心指导！我们已对全文括号内的引文进行了精简，仅保留最必要、

最具代表性的文献引用。仅在少数必须同时覆盖不同维度或单一文献无法充分支撑论点时，我们才保留了 2 条引文，且每处保留均经过仔细权衡，确保论证的严谨性。所有改动之处均已标橙，恳请您再次审阅！

……

审稿人 2 意见：

作者对前面提出的审稿意见作了认真细致的回答，稿件质量有较大提高，但仍存在如下一些主要问题。

意见 1：

该元分析试图揭示老年阶段启动的外语学习对个体执行功能的影响，由于相关研究的质量差异较大，质量最高的是“外语学习-主动对照”RCT 研究，建议敏感性分析部分针对“外语学习-主动对照”RCT 研究进行，以排除低质量研究对研究结论的影响。其次，主分析（g=0.37，显著）与敏感性分析（主动对照 RCT，g=0.11，不显著）结论相悖，这是本研究的核心发现。如果摘要只报显著结果，属于选择性报告（selective reporting），严重违背 PRISMA 规范。最后，相关结论的表述也需要谨慎，如采用“特异性语言增益有限，总体效应可能部分源于一般认知刺激和社会参与”或类似表述。

回应：

非常感谢审稿专家的耐心指导！针对上述建议，我们已做出如下修改：

首先，在研究结果部分**（3.3 主效应分析和异质性检验）的第一段**新增了针对“外语学习-主动对照”组随机对照试验（RCT）的敏感性分析，恳请您审阅该部分的标绿字体。实际上，这一段我们重新梳理了 4 个关键的效应量：1) 主分析总效应量，2) 所有 RCT 研究的总效应量，3) “外语学习-主动对照”的总效应量，4) RCT 研究中“外语学习-主动对照”的总效应量，恳请您再次审阅！

其次，我们已在**中英文摘要**和**“5 结论”**中同步补充了高质量“外语学习-主动对照”RCT 研究的结果（见相应处的标绿字体），确保报告符合 PRISMA 规范。

最后，我们对相关结论的表述也进行了修改。第一，在**“4.1 老年阶段外语学习对执行功能影响的主效应”的第一段**，我们按上述 4 个关键效应量进行了讨论（见该段标绿的 4 个句子），并进行了谨慎解释（见该段标红字体），确保结果呈现完整透明。第二，在“4.2.1 干预特征”的“(3) 对照类型”中修改了相关结论的表述（见该部分标绿字体）。

意见 2：

其实，老年阶段外语学习对两项执行功能（工作记忆、抑制控制）影响的效应量没有差异，加工复杂性效应假说的支持力度显著高于干扰抑制效应假说，以及支持双语执行加工优势假说而非双语抑制控制优势假说，这些结果都提示干预效果是综合性的，并非特异性指向抑制控制能力或执行功能。这与意见 1 中提到的关键点一致，即认知获益的来源是综合性的，不仅来源于语言特异性训练，也可能部分源于一般认知刺激或社会参与。建议作者在讨论部分增加相关内容的讨论。

回应：

非常感谢审稿专家的悉心指导！我们已在 **4.2 节末尾**新增一段讨论（见节末标绿字体），综合讨论了加工复杂性效应支持更强、执行功能成分无差异（支持 BEPA 而非 BICA）等结果，并指出这些发现与主动对照组的敏感性分析结果相互印证，共同表明外语学习的认知获益具有一定的综合性特征，可能部分源于“一般认知刺激或社会参与”。此外，本研究发现“外语学习-被动对照”与“外语学习-无干预”两类研究的效应量相同（g = 0.38），与原本的假设

“‘外语学习-无干预’效应量(综合效应量)最大”不符，说明社交参与对执行功能改善的独特贡献可能较小。因此，就本研究而言，外语学习对执行功能的总体积极改善主要源于“一般认知刺激”这一属性。这一结论也已在摘要、结论中写明。

此前碎片化呈现的结果经此修改更加清晰完整，由衷感谢审稿专家的认真审阅！

意见 3：

认知障碍老人的研究只有三项（5 个独立的效应量），研究数太少，建议如果把结果体现在摘要中，需要标注“初步证据显示（基于仅 3 项研究，需谨慎解读）”，或者摘要中完全不提认知障碍亚组，仅在正文结果与讨论部分详述，并明确标注“探索性分析，样本量不足，待后续验证”。

回应：

非常感谢审稿专家指出这一重要问题。考虑到汇报的完整性，我们选择在摘要中保留认知状态调节效应结果，同时根据您的建议补充了明确限定：标注该结果仅基于 3 项研究，样本量有限，需谨慎解读并待后续研究验证（见中英文摘要的标蓝字体）。同时也在讨论部分（**“4.2.2 受试特征”的“（2）认知状态”的第 2 段**）与“**5 结论**”这两处对样本局限性与探索属性进行了补充说明，请您查看相应段落的标蓝字体。

意见 4：

该元分析以标准化均值差 Hedges’*g* 为效应值，建议在“2.3.1 效应值计算”部分进一步交代标准化均值差（SMD）的计算方法，是“组内前测/后测对比（SMD）”、“组间后测对比（SMD）”还是“基线校正后的组间后测对比（SMD）”。

回应：

感谢审稿专家的指导！我们已在“2.3.1 效应值计算”部分进行了修订，说明了 Hedges’*g* 的计算方式，恳请您查看该节第一段的标蓝字体。

意见 5：

其他小问题：1）标题中“老年外语学习”建议修改为“老年阶段外语学习”；2）摘要中，“单变量分析”应改为“调节效应分析”；3）表述啰嗦，建议对引言部分进行缩减。

回应：

非常感谢审稿专家的细致指正，我们已逐条落实：

1）已将标题中的“老年外语学习”调整为“老年阶段外语学习”；

2）已在中英文摘要、结论部分将原表述“单变量分析”更正为“调节效应分析”(见相应处标红字体)；在结果与讨论部分，为与“多重回归分析”进行区分，保留了“单变量分析”这一表述；

3）已对引言部分进行精简，行文更凝练紧凑。

这几处细节调整对本研究表述的精准性提升帮助极大，再次感谢您的耐心指导！

第三轮

审稿人 1 意见：

作者对第二轮审稿意见作了认真细致的回答，文章内容和表述都有较大提高，同意发表。

编委 1 意见：

同意发表。

回应：

非常感谢编委对本稿件的发表推荐，我们由衷感激专家们为审阅稿件所付出的时间与专业指导。

编委 2 意见：

论文选题有意义，元分析程序清楚，并已经按照同行要求进行了修改，修改说明比较清楚。需要作者补充请提供关于异质性分析主要变量的编码细节，可以作为附件，供同行审议，包括干预特征：语言距离(近；远)、什么是近，什么是远？受试特征认知状态(认知健康；认知障碍)，按照什么标准？年龄的全距、平均数与标准差；如分组，则提供分组的标准、学习时长如何界定？全距与集中、离散分布；如分组，怎样分组？学习强度如何界定？全距与集中、离散分布；如分组，怎样分组？

回应：

非常感谢编委的宝贵意见！本研究中异质性分析所涉及的所有变量，此前已在“1.2 老年阶段外语学习与执行功能关系的影响因素”中进行了操作化定义，但分布相对零散，不利于同行审议。根据您的意见，我们将所有调节变量的**操作化定义和描述统计信息**进行了整合，统一放在附录附表 2 的末尾（见标红字体），恳请您再次审阅。

第四轮

编委 2 意见：

语言距离（分类变量）：目标语与受试母语同属一个语系编码为“近”；跨语系编码为“远”。编码如下：请补充，谁近，谁远。

回应：

非常感谢编委的细致审阅。针对语言距离的编码问题，我们新增“**附表 3 语言距离编码明细表**”，详细呈现各研究的受试母语、目标语及其语系归属和语言距离分类。并在“**附表 2 编码表**”后的说明性文字中增添了总体描述：

“本研究中，近距离语言对皆落在印欧语系内部；远距离语言对包括以下两种类型：（a）汉藏语系-印欧语系；（b）印欧语系-孤立语言。各效应量对应的具体语言组合及编码明细见附表 3。”

以上修改均已在修订稿中标绿，恳请编委再次审阅，感谢您的悉心指导！

第五轮

主编意见：

意见 1：

“纳入的 20 项研究”，只有 20 项研究，是否合适做元分析，是否合适做三水平元分析。如果不急，可以等 1~2 年后，文章数量多了之后再做元分析。

回应：

非常感谢主编的审慎建议！本研究最终纳入 20 项研究、63 个效应量，确实反映出目前

老年阶段外语学习干预领域的实证研究积累仍较为有限。鉴于部分研究同时报告多个效应量，效应量之间存在研究内依赖关系，因此本研究仍保留三水平元分析框架，以与现有数据结构相匹配。未来随着该领域实证研究数量增加，有必要及时更新元分析结果，以进一步检验本研究结论的稳定性。再次感谢主编的建议！

意见 2：

主效应：建议以表格的方式呈现。

回应：

非常感谢主编的宝贵建议！我们已在修改稿 3.3 节“主效应分析和异质性检验”中新增了表 1“老年阶段外语学习对执行功能影响的主效应及限定样本分析”，对主效应及相关限定样本分析结果进行了集中呈现。具体包括全部纳入研究、RCT 研究、“外语学习-主动对照”研究以及 RCT 研究中的“外语学习-主动对照”研究四类分析，并统一报告效应量数量、Hedges' *g*、95%置信区间及显著性水平，使主效应及其在不同限定样本下的结果更加直观清晰。与此同时，在 3.3 节正文开头增加“如表 1 所示”的引导语，使正文叙述与表格呈现相互对应。相关修改请见修改稿 3.3 节的标蓝文字。

意见 3：

表 1 有点让人费解，应该是针对主效应的结果进行调节效应检验？请作者再核实。

回应：

非常感谢主编的细致审阅！我们对此进行了核实。原表 1（现已修改为表 2）呈现的内容的确基于主效应分析发现总体效应存在显著异质性，进一步通过调节效应检验考察不同研究特征是否能够解释该主效应的异质性。原稿中这一分析逻辑交代得不够明确，且原表题仅为“调节效应检验”，未能清楚说明其所针对的具体效应，因此容易给读者带来困惑。

针对这一问题，我们进行了两方面修改。第一，在 3.4 节“调节效应检验”开头新增了“鉴于主效应分析显示总体效应存在显著异质性，进一步对可能影响老年阶段外语学习执行功能增益的特征进行调节效应检验”这一衔接说明，以明确调节效应检验是在前述主效应及异质性检验基础上开展的。第二，由于本轮新增了主效应结果表，原表 1 相应顺延为表 2，并将原题“调节效应检验”修改为“老年阶段外语学习对执行功能影响的调节效应检验”，进一步明确调节效应所针对的是外语学习对执行功能的影响。正文中相应的“如表 1 所示”也同步调整为“如表 2 所示”。相关修改请见修改稿 3.4 节的标蓝文字。