

Systematic Review

Phonological and Semantic Fluency in Alzheimer's Disease: A Systematic Review and Meta-Analysis

Rocio Olmos-Villaseñor^{a,b}, Consuelo Sepulveda-Silva^{a,b}, Teresa Julio-Ramos^b, Eduardo Fuentes-Lopez^a, David Toloza-Ramirez^c, Rodrigo A. Santibañez^{d,e}, David A. Copland^f and Carolina Mendez-Orellana^{a,b,*}

^a*Speech and Language Therapy School, Health Sciences Department, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile*

^b*Laboratory of Language Rehabilitation and Stimulation (LARES), Speech and Language Therapy School, Health Sciences Department, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile*

^c*Exercise and Rehabilitation Sciences Institute, School of Speech Therapy, Faculty of Rehabilitation Sciences, Universidad Andres Bello, Santiago, Chile*

^d*Neurology Department, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile*

^e*Neurology Service, Complejo Asistencial Doctor Sótero del Río, Santiago, Chile*

^f*Queensland Aphasia Research Centre, School of Health and Rehabilitation Sciences, University of Queensland, Brisbane, Australia*

Handling Associate Editor: Yasunori Yamada

Accepted 21 June 2023

Pre-press 15 July 2023

Abstract.

Background: Semantic and Phonological fluency (SF and PF) are routinely evaluated in patients with Alzheimer's disease (AD). There are disagreements in the literature regarding which fluency task is more affected while developing AD. Most studies focus on SF assessment, given its connection with the temporoparietal amnesic system. PF is less reported, it is related to working memory, which is also impaired in probable and diagnosed AD. Differentiating between performance on these tasks might be informative in early AD diagnosis, providing an accurate linguistic profile.

Objective: Compare SF and PF performance in healthy volunteers, volunteers with probable AD, and patients with AD diagnosis, considering the heterogeneity of age, gender, and educational level variables.

Methods: A total of 8 studies were included for meta-analysis, reaching a sample size of 1,270 individuals (568 patients diagnosed with AD, 340 with probable AD diagnosis, and 362 healthy volunteers).

Results: The three groups consistently performed better on SF than PF. When progressing to a diagnosis of AD, we observed a significant difference in SF and PF performance across our 3 groups of interest ($p=0.04$). The age variable explained a proportion of this difference in task performance across the groups, and as age increases, both tasks equally worsen.

*Correspondence to: Carolina Mendez-Orellana, PhD, Speech and Language Therapy School, Health Sciences Department, Faculty of Medicine, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Macul, Santiago, Chile. E-mail: carolinamendez@me.com.

Conclusion: The performance of SF and PF might play a differential role in early AD diagnosis. These tasks rely on partially different neural bases of language processing. They are thus worth exploring independently in diagnosing normal aging and its transition to pathological stages, including probable and diagnosed AD.

Keywords: Alzheimer's disease, phonemic fluency, semantic fluency, verbal fluency

INTRODUCTION

Clinical diagnosis of Alzheimer's disease (AD) includes impairment in two or more cognitive domains, such as memory, executive function, and behavior, which compromise social cognition [1], according to the standard criteria used, as determined by the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) [2].

Language is a domain of potential impairment that has been suggested as an important marker in assessing early AD stages, particularly relating to semantic fluency (SF) and phonological fluency (PF) [3, 4]. SF requires the ability to access and retrieve semantic knowledge, while PF depends on phonological/lexical retrieval mechanisms [5]. Studies focused on normal aging have highlighted that declines in lexical access affect performance on both tasks; however, impairment in executive function has also been linked with worse measures of SF and PF. Those findings mentioned how age could contribute to perseverative errors while performing these tasks [6–8]. SF is primarily relevant in dementia research due to its connection with the temporoparietal lobe amnesic system [9, 10]. Although PF is less reported, it is related to working memory, a cognitive function that is also impaired in patients with probable and confirmed diagnosis of AD [11, 12]. Given the functions these tasks target, both tasks are expected to be impaired in patients with dementia [11].

However, there are discrepancies regarding the performance of patients diagnosed with AD on SF and PF tasks [13]. This challenges the selection of which task is more critical for early diagnosis of patients with cognitive decline or dementia. Adlam et al. [14] studied SF performance in patients with AD diagnosis showing its impairment even at mild stages of AD. They also reported better performance on PF than SF, consistent with other studies [15–18]. In contrast, a retrospective study [19] comparing AD and frontotemporal dementia patients showed that the AD group performed better on SF than PF, consistent

with Comesaña and Coni [20], who found that AD patients showed more significant problems with PF than SF tasks.

Studies reporting normative data on SF and PF task performance are available in different countries, providing further insight on the protocols applied to evaluate fluency and the effects of demographic variables' effects when performing these tasks [21, 22]. Various protocols are used to study SF and PF, implementing different categories (e.g., generate animal names, supermarket items) or different amounts of letters. These tasks are generally evaluated in one minute [23, 24]; however, other study protocols consider other time windows, such as 90 seconds [25] or even no time limit to produce the words [26]. Task performance has been reported using raw results [27] or a sum of words produced for each letter or category [14, 28].

Two meta-analyses reporting SF and PF in healthy volunteers and AD patients [29, 30] include diverse study protocols evaluating SF and PF. Performance on SF and PF tasks depends critically upon the letter or semantic category chosen [31] and the time limit given. Age and educational level influence SF and PF task performance as well, and there are discrepancies in the literature on the brain regions for different verbal fluency tasks [32–35]. In patients with mild cognitive impairment and with AD, Kawano et al. [36] found that years of education were significantly related to SF test scores, but not significantly related to PF test scores. Similar effects of years of education and age have been consistently reported in many languages [22, 37, 38]. One reason that might be interfering in the comprehension of age and education level, is the variability of cognitive reserve in adult population [39]. Cognitive reserve refers to differences between individuals in how tasks are performed that might enable some people to be more resilient to brain changes than others. This “reserve” is combination of lifestyle choices and lifetime experiences, such as social activities and education [40]. A better understanding of the concept of cognitive reserve could lead to interventions to mitigate cognitive ageing or reduce the risk of dementia [39, 41, 42]. Although less reported, there are also conflicting

findings regarding whether gender affects SF and PF performance [21, 22, 36].

So far, no previous meta-analyses have considered including the same parameters (time, category, letter, score) for evaluating SF and PF performance in this clinical population. The processing behind these fluency tasks might reveal better insight into the common word-finding problems in healthy and pathological aging. The present systematic review and meta-analysis thus aimed to compare SF and PF performance in healthy volunteers, and in individuals with probable and confirmed AD diagnoses, considering the influence of age, gender, and educational level.

METHODS

Review design

A systematic review and meta-analysis were conducted following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [43]. The manuscript was registered in PROSPERO under the code CDR42022352253. The PRISMA checklist appears in the Supplementary Material.

Data sources and search strategy

We reviewed the literature in PubMed, EMBASE, and CENTRAL databases, from January 1975 to December 2020. Our search strategy was performed in each database, selecting relevant terms aligned with our research question. We used the terms “Alzheimer’s Disease, Alzheimer’s type dementia, aged” to characterize our interest population. We selected terms related to performance in the SF and PF task, such as “Verbal Fluency, Verbal Fluency Test, Semantic Fluency, and Phonemic Fluency.” All synonyms of relevant concepts were included and were linked with Boolean terms such as AND and OR. We used controlled vocabulary assigned by the indexers, Mesh for PubMed and Central, Emtree for the Embase database, and natural language. The complete search strategy and filters used for the three databases are described in Supplementary Tables 1–3.

Eligibility criteria

Inclusion criteria considered studies reporting: a) adults with probable diagnosis or diagnosed with AD according to NINCDS-ADRDA criteria, b) report on

both tasks, SF and PF, c) SF and PF tasks assessed following the standard time of one minute, d) SF evaluation with “animal” category according to varying national norms, e) PF evaluation with validated letters according to each country with raw score, f) articles in English and Spanish, and g) studies reporting baseline measurements. We excluded case studies, case analyses, and systematic reviews (with or without meta-analysis).

Data collection process

Duplicated articles were identified using the features included in the Rayyan software [44]. Two authors (R.O.V. and C.S.S.) applied inclusion and exclusion criteria to all titles and abstracts, selecting studies according to the inclusion criteria. The interrater reliability (IRR) determined with Cohen’s K was 0.69 ($p < 0.001$). The 10% of disagreements were resolved by a third senior reviewer (C.M.O.). The authors applied their self-designed checklist to extract the included articles’ data and synthesize the evidence (Supplementary Table 4).

Risk of bias assessment

The risk of bias and the applicability of the eight full-text articles included in this systematic review and meta-analysis was performed by the QUADAS-2 tool [45]. QUADAS-2 is an instrument designed and validated to assess bias risk and the applicability of diagnostic accuracy studies. Two reviewers independently graded each article (R.O.V. and C.S.S.), and the scores were compared. A third reviewer (C.M.O.) resolved disagreements between the two principal reviewers.

Statistical analysis

A systematic review and meta-analysis were performed for the reported SF and PF scores among the three groups, and the weighted average was estimated. The values per task were considered to be the number of words generated in one minute, and an unstandardized mean difference was estimated for the number of words. We used Review Manager 5.4, and with Cochran’s Q statistic test, we analyzed the differences between SF and PF across all groups, comparing the means of each group extracted across studies. Due to the highly heterogeneous values reported in the selected studies (>75%), we performed a random-effects meta-regression including

the following demographic variables: age, educational level, and gender (proportion of females) in each study. We estimated the percentage of heterogeneity R-squared (%) attributable to each demographic variable. Univariate and multivariate meta-regression models were created using the “meta regress” command of STATA v16 software, applying the Knapp-Hartung method with truncation to adjust the standard error of coefficients. Multivariate models were constructed with the variables that showed significant association with the univariate models. The proportion of variance explained between stud-

ies (R^2) came from the univariate and multivariate meta-regression models.

RESULTS

Study selection

Figure 1 shows the flow chart for the selection process based on the PRISMA guidelines [43]. A total of 5,902 potential studies were identified. After removing duplicated articles (1,160), 4,742 studies were considered for screening titles and abstracts, exclud-

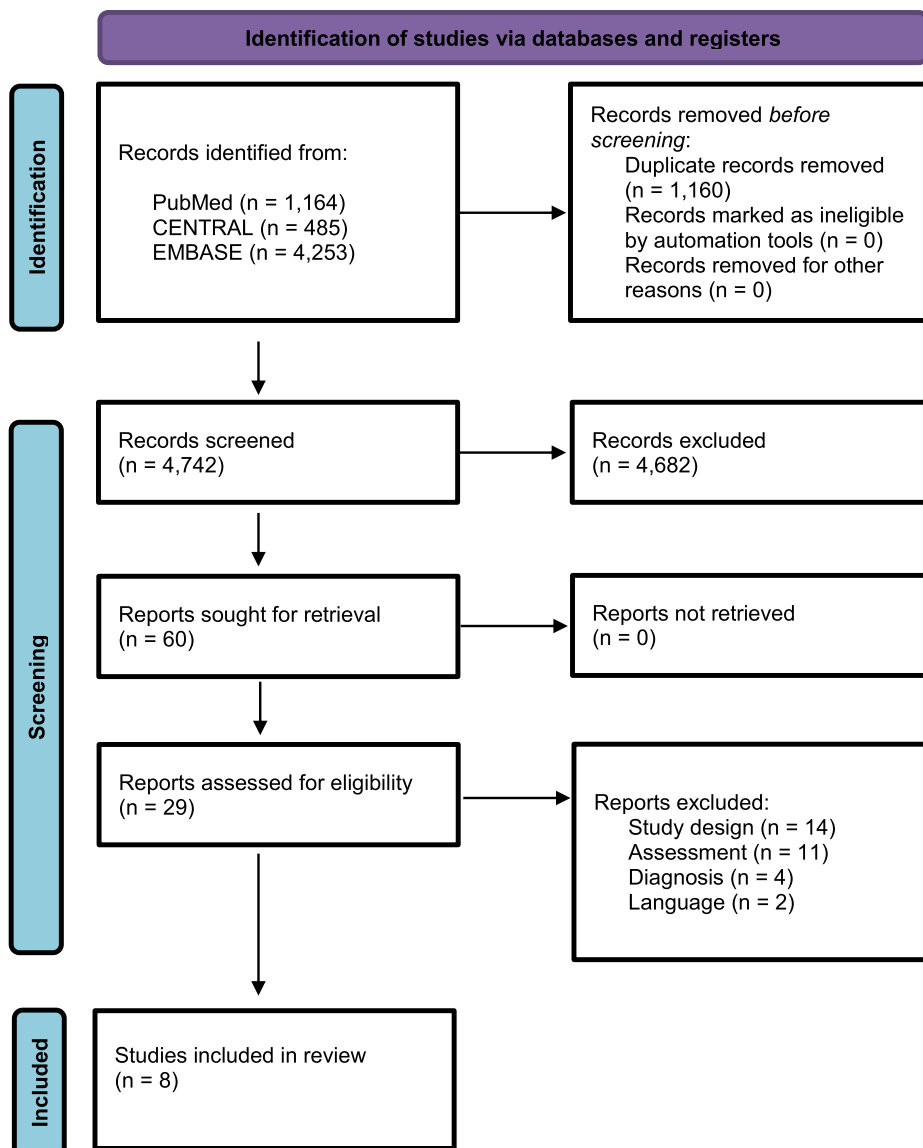


Fig. 1. Flow Diagram PRISMA search and selection articles.

ing 4,682 articles. After screening, 60 articles were considered for critical analysis. After applying inclusion criteria, 29 studies were selected, but we could extract data from only 8 studies for the meta-analysis of both SF and PF [12, 22, 27, 36, 46–49].

Study characteristics

Table 1 shows the general characteristics of the eight articles included in the review. The sample size of the selected studies was 1,270 patients (568 patients diagnosed with AD (44.7%), 340 volunteers with a probable AD diagnosis (26.7%), and 362 healthy volunteers (28.5%)), and the average age of all groups was 71.5 years. A total of 738 participants (58%) were women, while one article did not report the number of participants in their study. Education level showed an average of 10.1 years of schooling in the total sample. Finally, Table 1 shows the heterogeneity of years of publication (the oldest published in 1999), and the country for all the articles included.

Bias risk in the selected studies

Using the QUADAS-2 tool [45], all the articles were analyzed considering risks of bias and applicability (Fig. 2). Regarding the risk of bias for Domains 1 and 2 (patient selection and index test, respectively), 90% were classified as low risk. In addition, in Domains 3 and 4 (reference standard and flow and timing, respectively), 80% of the articles were classified as low risk. On the other hand, the assessment of applicability concerns showed that 90% were classified as low risk for patient selection and index tests. By contrast, 80% of the articles were classified as low risk and 20% as high risk for reference standard evaluation.

Results of syntheses

Performance on SF and FP tasks was compared between the three groups: healthy volunteers, volunteers with a probable AD diagnosis, and patients diagnosed with AD. Although significant heterogeneity was reported across studies ($I^2 = 90.5\%$), Cochran's Q test showed significant differences between both types of verbal fluency among all groups ($p = 0.04$). The lower difference in SF and PF task performance corresponds to patients diagnosed with AD (mean difference 2.93; 95%CI:1.59–4.27). Figure 3 shows SF and PF performance in the three groups. In the intra-group analysis for the

healthy volunteers, we observed high heterogeneity ($I^2 = 80.20\%$) across the values reported in the selected studies. However, the difference in task performance between both tasks was statistically significant ($p < 0.001$). Healthy volunteers perform better for the SF, generating approximately 6 words more than PF compared with the other two groups. The studies included for analyzing volunteers with a probable diagnosis of AD also demonstrated high heterogeneity ($I^2 = 90.58\%$). The difference in task performance remained significant, though ($p < 0.001$), with higher scores for SF, approximately 4 words more than for PF. The studies included for the diagnosed AD patient group showed high heterogeneity ($I^2 = 88.10\%$), and the difference between SF and PF tasks was statistically significant ($p < 0.001$). This group also performed better in SF, generating approximately 3 words more than for the PF task.

The univariate model built with the variable group showed high heterogeneity ($I^2 = 86.67\%$), but task performance remained significantly different ($p = 0.025$), being explained by the differences between groups ($R^2 = 24.04\%$). The group of patients diagnosed with AD generated approximately 3 words less than the healthy volunteer group ($\beta = -2.74$; 95% CI: -5.09 to -0.40 ; $p = 0.025$). The univariate model built with the variable age showed high heterogeneity ($I^2 = 83.83\%$), and the tasks performance difference was significant ($p = 0.01$). The age variable explains task performance differences ($R^2 = 41.06\%$) with a negative coefficient of -0.18 (95% CI -0.33 to -0.05 ; $p = 0.011$). Age influenced task performance, with fewer words generated as age rose. The univariate model built with the variable gender (considered as the proportion of females) showed high heterogeneity ($I^2 = 89.27\%$) but did not reach significance in explaining the varying heterogeneity of the task performance ($R^2 = 9.27\%$; $p = 0.1$). The univariate model built with educational level was also highly heterogeneous ($I^2 = 90.47\%$) but did not reach significance explaining the heterogeneity in task performance differences ($R^2 = 0.00\%$; $p = 0.319$).

A multivariate model built considering the group and age variable revealed that a large proportion of task performance difference could be explained by the variable age ($R^2 = 43.60\%$) besides its high heterogeneity ($I^2 = 82.24\%$). However, age was not statistically significant ($p = 0.06$), explaining the task performance. A lack of statistical power could explain the non-significance of this variable (see below). The negative coefficient (-0.13) demonstrates that task

Table 1
General characteristics of the articles

Source	First author	Year	Type of study	Country	Target group	Educational level mean (y)	Age mean (y)	Sex	
								♀ n	♂ n
Verbal fluency in Alzheimer's disease and mild cognitive impairment in individuals with low educational level and its relationship with reading and writing habits [13].	Tessaro et al.	2020	Controlled clinical trial	Brazil	Total sample: 43 Healthy adults (n = 25), AD group (n = 18)	Healthy adults 4.84 AD group 5.28	Healthy adults 67.76 AD group 73.89	–	–
Normative data for phonemic and semantic verbal fluency test in the adult French-Quebec population and validation study in Alzheimer's disease and depression [22].	St-Hilaire et al.	2016	Controlled clinical trial	Canada	Total sample: 124 Healthy adults (n = 62), AD group (n = 62)	Healthy adults 14.4 AD group 11.8	Healthy adults 70.4 AD group 76.8	81	43
Word recognition in Alzheimer's disease: Effects of semantic degeneration [46].	Cuetos et al.	2017	Controlled clinical trial	UK	Total sample: 65 Healthy adults (n = 25), AD group (n = 40)	Healthy adults 8.16 AD group 8.1	Healthy adults 77.2 AD group 78.7	56	9
Category and letter fluency in semantic dementia, primary progressive aphasia, and Alzheimer's disease [47].	Marczinski et al.	2006	Controlled clinical trial	Canada	Total sample: 40 Healthy adults (n = 20), AD group (n = 20)	Healthy adults 13.5 AD group 9.9	Healthy adults 73.8 AD group 74.7	20	20
Heterogeneity in executive impairment in patients with very mild Alzheimer's disease [27].	Stokholm et al.	2006	Controlled clinical trial	Denmark	Total sample: 68 Healthy adults (n = 32), AD group (n = 36)	Healthy adults 11.6 AD group 10.7	Healthy adults 74.3 AD group 76	40	28
Mirror asymmetry of Category and Letter fluency in traumatic brain injury and Alzheimer's disease patients [48].	Capitani et al.	2009	Prospective study	Italy	Total sample: 255 Healthy adults (n = 198), AD group (n = 57)	Healthy adults 10.8 AD group 6.81	Healthy adults 48.7 AD group 71.95	109	146
Prevalence and correlates of category versus letter fluency discrepancies in Alzheimer's disease [49].	Sherman et al.	1999	Prevalence study	USA	Total sample: 217 Probable AD group (n = 217)	13.3	72.6	144	73
Effects of educational background on verbal fluency task performance in older adults with Alzheimer's disease and mild cognitive impairment [36].	Kawano et al.	2010	Controlled clinical trial	Japan	Total sample: 458 AD group 1, (n = 125), AD group 2, (n = 210), Probable AD group 1 (n = 39), Probable AD group 2 (n = 84)	AD group 1 9 AD group 2 12 Probable AD group 1 9 Probable AD group 2 12	AD group 1 70.06 AD group 2 69.66 Probable AD group 1 69.96 Probable AD group 2 69.02	288	170

–, not reported; UK, United Kingdom; USA, United States of America.

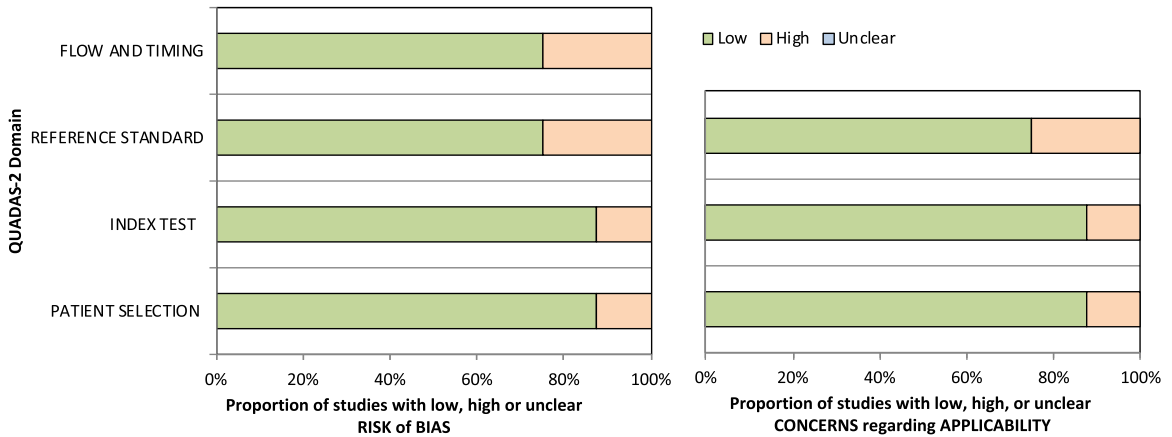
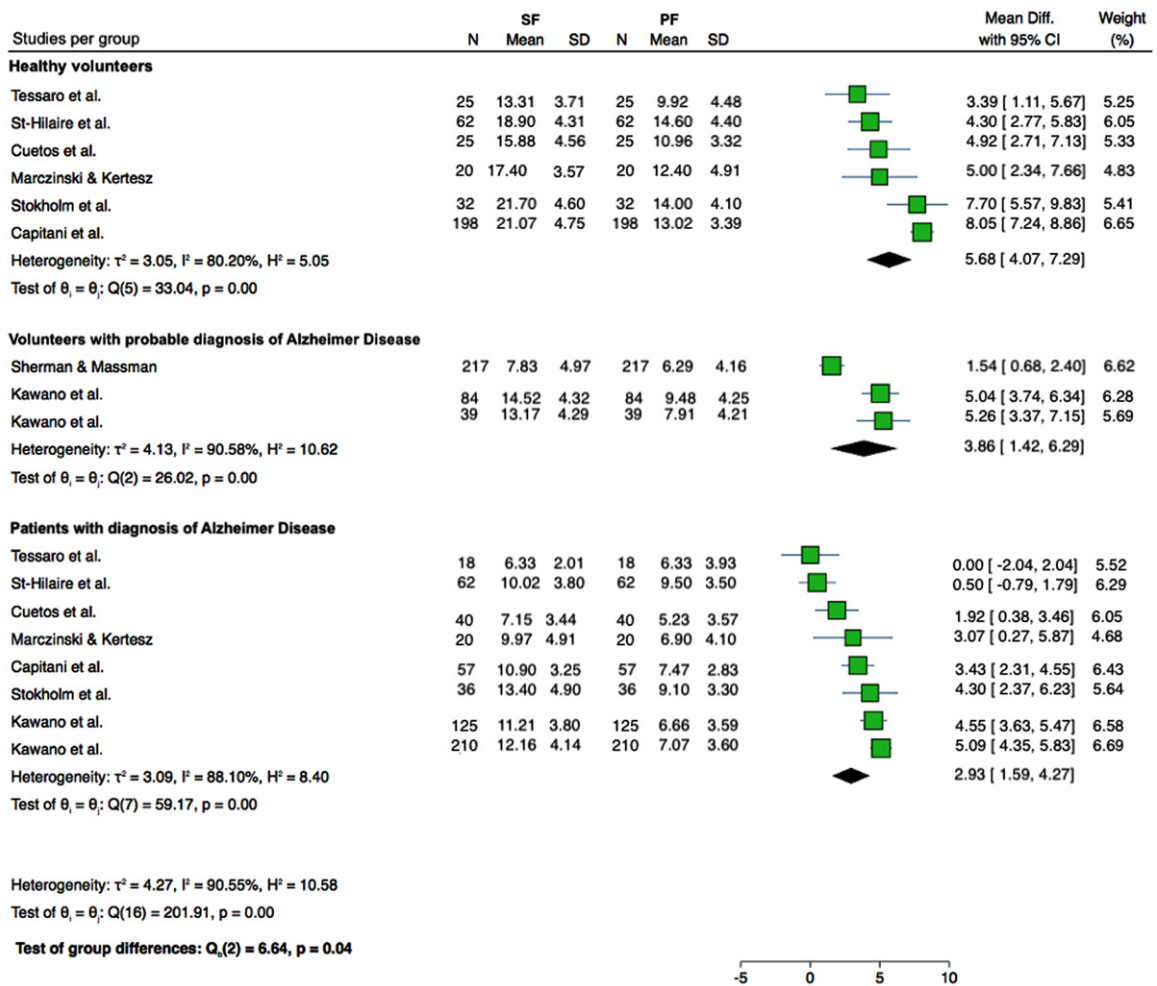


Fig. 2. Assessment of the risk of bias and applicability for the reviewed studies.



Random-effects REML model

Random-effects REML model. Results correspond to raw scores of SF and PF.

Fig. 3. Analysis of the performance during SF and PF in the three groups of interest. Random-effects REML model. Results correspond to raw scores of SF and PF.

performance difference is affected by increasing age in each group.

DISCUSSION

The present study aimed to compare SF and PF task performance in three groups of interest: healthy volunteers, volunteers with a probable AD diagnosis, and diagnosed AD patients, considering the variables age, gender, and years of education. Applying diverse protocols to measure SF and PF results in a different valorization of task performance. We thus implemented strict inclusion criteria to select studies that reported the same protocol to evaluate SF and PF in healthy adults, adults with a probable AD diagnosis, and diagnosed AD patients. Our results support the claim that performance in SF and PF tasks significantly differs in healthy volunteers. We observed that the three groups consistently performed better on SF than PF. However, when progressing to an AD diagnosis, SF and PF performance deteriorates equally, but significant differences in task performance remain. This finding challenges the practice of focusing primarily on semantic processing in diagnosing dementia, highlighting the benefits of examining both tasks' performance from the early stages.

We found that healthy volunteers produced more words in the SF task than patients with suspected or confirmed AD diagnoses, similar to the findings of Saranpää et al. [50] in patients with mild cognitive impairment and AD patients. However, we observed this pattern not only for SF but also for PF. Although high heterogeneity was reported across our selected studies, Cochran's Q test showed significant differences between both types of fluency tasks among all groups ($p=0.04$). This difference is pronounced in the group of healthy volunteers and in individuals where the disease has progressed; performance in both tasks decreased, but these differences remained significant, with SF performance better than PF performance. Our results are aligned with studies that postulate that SF is less impaired than PF in healthy and pathological aging [19, 20]. Hart et al. [31] found a similar result in a sample of subjects with AD, contradicting the prevailing hypothesis that SF is the most affected in this pathology. Later, Comesaña and Coni [20] confirmed this finding, reporting that diagnosed AD patients show better SF performance than PF [20]. A retrospective study also showed that AD performed better in SF tasks [19]. A sample of patients

with mild to moderate AD found retrieving words from semantically defined categories easier than by first letters [51]. While there is significant evidence that SF performance is more impaired in healthy subjects as they progress to AD stages [15, 52], our results report the opposite.

SF and PF performance depends on partially different cognitive-linguistic systems, and the deterioration of these systems is different as AD progresses. PF requires a greater demand of frontal lobe mediated cognitive processes [53], imposes fast mental processing [54], information evocation, and search, and requires decoding as well as encoding, specifically linked to the initial letter [55]. On the other hand, SF requires retrieval strategies that use the meaning of words within a semantic category and generate consecutive responses that are semantically similar. This occurs partly because of automatic associative retrieval processes [56] drawing inferences about the relatedness of concepts encoded in semantic memory [57]. Several studies argue that as language is semantically represented and organized [58, 59], retrieving words from a semantic category would be consistent with how language is stored in the mind. In contrast, word representations are not alphabetically organized [60], meaning that PF requires engaging additional cognitive strategies. Letter retrieval involves exploring more subsets of categories than retrieving names from a specific semantic category [61, 62], which could be related to the worse performance in PF found in our findings, given possible higher task demands in PF compared to SF. A previous meta-analysis investigated healthy volunteers' brain activation during phonemic and semantic verbal fluency tasks. Evidence arose for spatially different activation in BA 44, but not in other regions of the left inferior frontal gyrus and the middle frontal gyrus (BA 9, 45, 47) during phonemic and semantic verbal fluency processing [63]. Schmidt and colleagues [64] also identified circumscribed parts of the left inferior frontal gyrus and left superior and middle temporal gyrus that are significantly double dissociated concerning their differential contribution to PF and SF, respectively. Meinzer et al. [65] reported differential brain activity patterns of older and younger adults performing SF and PF tasks. Young adults recruited different subparts of the left inferior frontal gyrus for SF and PF, but older adults did not show this distinction. The PF task was comparable for both age groups and was reflected in strongly left-lateralized (frontal) activity patterns. In older adults, poor PF performance was accompanied by additional right frontal activity,

while SF is more related to the temporal cortex function [66]. Merging behavioral data with task-driven neuroimaging data might shed light on the neural mechanisms underlying these tasks in healthy and pathological aging. Behind this differential performance in SF and PF, a neural network might start failing, and it might deteriorate faster in the presence of dementia. Based on SF and PF, early detection of deterioration and cognitive stimulation promotion during the early stages of aging could be done.

A univariate model was performed comparing the performance of SF and PF tasks between the groups of healthy volunteers, volunteers with a probable AD diagnosis, and diagnosed AD patients. Despite no significant differences, the volunteers with a probable AD diagnosis produced an average of 2 words less than the healthy volunteers in both fluency tasks ($p=0.2$; coefficient = -1.8). However, in the group of diagnosed AD patients, performance is significantly worse than the healthy volunteers ($p=0.02$; coefficient = -3), generating approximately 3 fewer words in each fluency task. For those with suspected AD and when the disease was diagnosed, there was an equivalent drop in performance for both fluency tasks as well. This decline in fluency performance has been associated with impaired executive functioning in healthy subjects and people with AD [12], and happens because both fluency tests require organization, inhibition, cognitive flexibility, and working memory, which are key executive components for word search and retrieval.

In our analyses, we adjusted for possible confounding factors, including demographic variables, such as age, years of education, and gender (considered as the proportion of females). Among these, only age significantly affected SF and PF task performance. Our univariate model results revealed that performance declines in both tasks as we age. This finding aligns with Olabarrieta-Landa et al. [21] and Pakhomov et al. [67]. By contrast, years of education failed to account for the difference in performance between verbal fluency tasks across the groups. The literature reports an association between years of education and better fluency performance [21, 36]; however, this variable is not consistently considered during clinical assessment. Our results revealed that performance declines throughout the establishment of the disease is not explained by years of education. This observation meshes with Sherman et al. [49], where no significant differences were found for age, education level, symptoms duration, or rate of dementia progression. Itaguchi et al. [68] also found that age and

education did not influence SF differences between healthy older and AD groups. On the other hand, the reported impact of gender on PF and SF performance is inconsistent or nonexistent in the literature [69]. The results of a study in Latin America suggest not adjusting for gender when obtaining percentiles for verbal fluency tests [21]. The study of St-Hilarie et al. [22] also observed that gender did not affect performance in each condition. However, Irvine et al. [42] found that cognitive functions are more severely and more widely affected in women with AD than in men, even accounting for differences in age, education, or dementia severity. Our analyses highlight that performance on fluency tasks is an indicator to be observed more routinely in clinical practice, which may be associated with AD diagnosis confirmation, since the decline in this performance is not explained by the variable of years of schooling or gender.

When exploring the age variable and groups of interest with a multivariate model, both variables fail to explain the difference in performance between verbal fluency tasks; however, we observed a trend for the age variable ($p=0.06$). Given that our meta-analysis used a truncated Knapp-Hartung method to estimate the standard error, we highlight this borderline finding for age. This method has much more appropriate false positive rates than the standard T-test for regression coefficients in meta-regression [70]; however, it might have affected the statistical power to detect the effect of the age variable in our groups.

This study has several limitations. We performed an extensive search, but due to our restricted inclusion criteria, we ended up with few studies that could provide the information needed for our meta-analysis. The articles excluded in this study showed different study designs and paradigms for assessing SF and PF tasks. Many of those studies also did not accurately report the early AD stage and did not report the diagnosis quantitatively. Extracting the raw results of the PF tasks was not always possible either, because some studies grouped and presented as a sum of the performance of the fluency test called "FAS" (fluency of three letters: F, A, and S). We did not consider patients diagnosed with AD with comorbidities in the inclusion criteria, which might also limit the extrapolation of our results. Finally, our review was conducted during the Coronavirus pandemic, and we included studies up to December 2020 thanks to the opportunity of open access initiatives and databases that have supported low- and mid-income countries.

Comparing SF and PF performance in volunteers with a probable AD diagnosis, patients diagnosed with AD, and healthy volunteers is of significant interest in research and clinical practice. Increasing knowledge about verbal fluency in these groups could help determine what verbal fluency tasks are essential to improve early detection and diagnosis of AD patients. Our results are relevant for future research because they highlight the importance of language as a functional cognitive domain that could establish differences between the transition from normal aging to pathological stages such as AD. One strength of this meta-analysis is that we considered studies published in both languages Spanish and English. Finally, an essential projection is the unification of protocols for assessing verbal fluency tasks (SF and PF). Words produced for semantic categories and a given letter thus have to be based on idiomatic heterogeneity, and it would be helpful to consider a normative letter for language and sociocultural characteristics in PF. The differences in performance through aging are of great interest to evaluate in combination with neuroimaging data, providing further understanding of the neural and cognitive changes associated with underlying pathological aging.

Conclusion

Performance on SF and PF tasks might play a differential role in early AD diagnosis. These two tasks target partially different neurocognitive mechanisms and should be explored for differential diagnosis between normal aging and its transition to pathological stages such as probable and diagnosed AD.

ACKNOWLEDGMENTS

This work was supported by the Health Science Department of Pontificia Universidad Católica de Chile, ANID-Subdirección de Capital Humano/Doctorado Nacional/2021-21212181 (David Toloza-Ramirez), and ANID-Subdirección de Capital Humano/Doctorado Nacional/2023-21230591 (Teresa Julio-Ramos).

FUNDING

The authors have no funding to report.

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

DATA AVAILABILITY

The data supporting the findings of this study are available within the article and/or its supplementary material. Additional information is available on request from the corresponding author.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: <https://dx.doi.org/10.3233/JAD-221272>.

REFERENCES

- [1] Begali VL (2020) Neuropsychology and the dementia spectrum: Differential diagnosis, clinical management, and forensic utility. *Neurorehabilitation* **46**, 181-194.
- [2] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939-939.
- [3] Blair M, Marczyński CA, Davis-Faroque N, Kertesz A (2007) A longitudinal study of language decline in Alzheimer's disease and frontotemporal dementia. *J Int Neuropsychol Soc* **13**, 237-245.
- [4] Toloza-Ramírez D, Méndez-Orellana C, Martella D (2021) Diagnóstico neuropsicológico diferencial en enfermedad de Alzheimer y demencia frontotemporal: Una revisión sistemática cualitativa. *Neurol Perspect* **1**, 82-97.
- [5] Butters N, Granholm E, Salmon DP, Grant I, Wolfe J (1987) Episodic and semantic memory: A comparison of amnesic and demented patients. *J Clin Exp Neuropsychol* **9**, 479-497.
- [6] Henry JD, Phillips LH (2006) Covariates of production and perseveration on tests of phonemic, semantic and alternating fluency in normal aging. *Aging Neuropsychol Cogn* **13**, 529-551.
- [7] Sauzéon H, Raboutet C, Rodrigues J, Langevin S, Schelstraete MA, Feyereisen P, Hupet M, N'Kaoua B (2011) Verbal knowledge as a compensation determinant of adult age differences in verbal fluency tasks over time. *J Adult Dev* **18**, 144-154.
- [8] Shao Z, Janse E, Visser K, Meyer AS (2014) What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Front Psychol* **5**, 772.
- [9] Wright LM, Marco MD, Venneri A (2022) Verbal fluency discrepancies as a marker of the prehippocampal stages of Alzheimer's disease. *Neuropsychology*. doi: 10.1037/neu0000836.
- [10] Delgado-Álvarez A, Cabrera-Martín MN, Pytel V, Delgado-Alonso C, Matías-Guiu J, Matias-Guiu JA (2021) Design and verbal fluency in Alzheimer's disease and frontotemporal dementia: Clinical and metabolic correlates. *J Int Neuropsychol Soc* **28**, 947-962.
- [11] Epker MO, Lacritz LH, Cullum CM (1999) Comparative analysis of qualitative verbal fluency performance in normal elderly and demented populations. *J Clin Exp Neuropsychol* **21**, 425-434.
- [12] Torres AL (2020) Biomarcadores de neuroimagen para el diagnóstico temprano de la enfermedad de Alzheimer. Un

- abordaje desde las redes neuronales. *Rev Chil Radiol* **26**, 105-112.
- [13] Tessaro B, Hermes-Pereira A, Schilling LP, Fonseca RP, Kochhann R, Hübner LC (2020) Verbal fluency in Alzheimer's disease and mild cognitive impairment in individuals with low educational level and its relationship with reading and writing habits. *Dementia Neuropsychol* **14**, 300-307.
- [14] Adlam A-LR, Bozeat S, Arnold R, Watson P, Hodges JR (2006) Semantic knowledge in mild cognitive impairment and mild Alzheimer's disease. *Cortex* **42**, 675-684.
- [15] Henry JD, Crawford JR, Phillips LH (2004) Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia* **42**, 1212-1222.
- [16] Murphy KJ, Rich JB, Troyer AK (2006) Verbal fluency patterns in amnesic mild cognitive impairment are characteristic of Alzheimer's type dementia. *J Int Neuropsychol Soc* **12**, 570-574.
- [17] Sawyer RP, Rodriguez-Porcel F, Hagen M, Shatz R, Espay AJ (2017) Diagnosing the frontal variant of Alzheimer's disease: A clinician's yellow brick road. *J Clin Mov Disord* **4**, 2.
- [18] Verma M, Howard RJ (2012) Semantic memory and language dysfunction in early Alzheimer's disease: A review. *Int J Geriatr Psychiatry* **27**, 1209-1217.
- [19] Reul S, Lohmann H, Wiendl H, Duning T, Johnen A (2017) Can cognitive assessment really discriminate early stages of Alzheimer's and behavioural variant frontotemporal dementia at initial clinical presentation? *Alzheimers Res Ther* **9**, 61.
- [20] Comesaña A, Coni AG (2013) Análisis de tareas de fluidez verbal en enfermos de Alzheimer y adultos sanos. *Temas Psicol* **21**, 269-281.
- [21] Olabarrieta-Landa L, Rivera D, Galarza-del-Angel J, Garza M, Saracho C, Rodríguez W, Chávez-Oliveros M, Rábago B, Leibach G, Schebela S, Martínez C, Luna M, Longoni M, Ocampo-Barba N, Rodríguez G, Aliaga A, Esenarro L, Cadena CG de la, Perrin B, Arango-Lasprilla J (2015) Verbal fluency tests: Normative data for the Latin American Spanish speaking adult population. *Neurorehabilitation* **37**, 515-561.
- [22] St-Hilaire A, Hudon C, Vallet GT, Bherer L, Lussier M, Gagnon J-F, Simard M, Gosselin N, Escudier F, Rouleau I, Macoir J (2016) Normative data for phonemic and semantic verbal fluency test in the adult French-Quebec population and validation study in Alzheimer's disease and depression. *Clin Neuropsychol* **30**, 1126-1150.
- [23] Kertesz A (2022) The Western Aphasia Battery: A systematic review of research and clinical applications. *Aphasiology* **36**, 21-50.
- [24] Dubois B, Slachevsky A, Litvan I, Pillon B (2000) The FAB: A Frontal Assessment Battery at bedside. *Neurology* **55**, 1621-1626.
- [25] Ober BA, Dronkers NF, Koss E, Delis DC, Friedland RP (1986) Retrieval from semantic memory in Alzheimer-type dementia. *J Clin Exp Neuropsychol* **8**, 75-92.
- [26] Hough MS, Givens G (2004) Word fluency skills in dementia of the Alzheimer's type for common and goal-directed categories. *Aphasiology* **18**, 357-372.
- [27] Stokholm J, Vogel A, Gade A, Waldemar G (2006) Heterogeneity in executive impairment in patients with very mild Alzheimer's disease. *Dement Geriatr Cogn* **22**, 54-59.
- [28] Sailor KM, Zimmerman ME, Sanders AE (2011) Differential impacts of age of acquisition on letter and semantic fluency in Alzheimer's disease patients and healthy older adults. *Q J Exp Psychol (Hove)* **64**, 2383-2391.
- [29] Henry JD, Crawford JR (2004) A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology* **18**, 284-295.
- [30] Laws KR, Duncan A, Gale TM (2010) 'Normal' semantic-phonemic fluency discrepancy in Alzheimer's disease? A meta-analytic study. *Cortex* **46**, 595-601.
- [31] Hart S, Smith CM, Swash M (1988) Word fluency in patients with early dementia of Alzheimer type. *Br J Clin Psychol* **27**, 115-124.
- [32] Abrahams S, Goldstein LH, Simmons A, Brammer MJ, Williams SCR, Giampietro VP, Andrew CM, Leigh PN (2003) Functional magnetic resonance imaging of verbal fluency and confrontation naming using compressed image acquisition to permit overt responses. *Hum Brain Mapp* **20**, 29-40.
- [33] Gordon JK, Young M, Garcia C (2018) Why do older adults have difficulty with semantic fluency? *Aging Neuropsychol Cogn* **25**, 803-828.
- [34] Hennessee JP, Reggente N, Cohen MS, Rissman J, Castel AD, Knowlton BJ (2019) White matter integrity in brain structures supporting semantic processing is associated with value-directed remembering in older adults. *Neuropsychologia* **129**, 246-254.
- [35] Ouyang M, Cai X, Zhang Q (2020) Aging effects on phonological and semantic priming in the tip-of-the-tongue: Evidence from a two-step approach. *Front Psychol* **11**, 338.
- [36] Kawano N, Umegaki H, Suzuki Y, Yamamoto S, Mogi N, Iguchi A (2010) Effects of educational background on verbal fluency task performance in older adults with Alzheimer's disease and mild cognitive impairment. *Int Psychogeriatr* **22**, 995-1002.
- [37] Aki OE, Alkan B, Demirsoz T, Velibasoglu B, Tasdemir T, Erbas SP, Selvi K, Ergenc I, Bariskin E, Ozdemir P, Demir B (2021) Effects of age, gender and education on phonemic and semantic verbal fluency. *Turkish J Psychiatry* **33**, 53-64.
- [38] Troyer AK, Moscovitch M, Winocur G (1997) Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology* **11**, 138-146.
- [39] Stern Y (2012) Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol* **11**, 1006-1012.
- [40] Oosterhuis EJ, Slade K, May PJC, Nuttall HE (2022) Toward an understanding of healthy cognitive aging: The importance of lifestyle in cognitive reserve and the scaffolding theory of aging and cognition. *J Gerontol B Psychol Sci Soc Sci* **78**, 777-788.
- [41] Reuter-Lorenz PA, Park DC (2014) How does it STAC up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychol Rev* **24**, 355-370.
- [42] Irvine K, Laws KR, Gale TM, Kondel TK (2012) Greater cognitive deterioration in women than men with Alzheimer's disease: A meta analysis. *J Clin Exp Neuropsychol* **34**, 989-998.
- [43] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Syst Rev* **10**, 89.
- [44] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A (2016) Rayyan—a web and mobile app for systematic reviews. *Syst Rev* **5**, 210.

- [45] Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM, QUADAS-2 Group (2011) QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* **155**, 529-536.
- [46] Cuetos F, Arce N, Martínez C, Ellis AW (2017) Word recognition in Alzheimer's disease: Effects of semantic degeneration. *J Neuropsychol* **11**, 26-39.
- [47] Marczyński CA, Kertesz A (2006) Category and letter fluency in semantic dementia, primary progressive aphasia, and Alzheimer's disease. *Brain Lang* **97**, 258-265.
- [48] Capitani E, Rosci C, Saetti MC, Laiacona M (2009) Mirror asymmetry of Category and Letter fluency in traumatic brain injury and Alzheimer's patients. *Neuropsychologia* **47**, 423-429.
- [49] Sherman AM, Massman PJ (1999) Prevalence and correlates of category versus letter fluency discrepancies in Alzheimer's disease. *Arch Clin Neuropsychol* **14**, 411-418.
- [50] Saranpää AM, Kivisaari SL, Salmelin R, Krumm S (2022) Moving in semantic space in prodromal and very early Alzheimer's disease: An item-level characterization of the semantic fluency task. *Front Psychol* **13**, 777656.
- [51] Rosen WG (1980) Verbal fluency in aging and dementia. *J Clin Neuropsychol* **2**, 135-146.
- [52] Clark LJ, Gatz M, Zheng L, Chen Y-L, McCleary C, Mack WJ (2009) Longitudinal verbal fluency in normal aging, preclinical, and prevalent Alzheimer's disease. *Am J Alzheimers Dis Other Dement* **24**, 461-468.
- [53] Arán Filippetti V (2011) Fluidez verbal según tipo de tarea, intervalo de tiempo y estrato socioeconómico, en niños escolarizados. *An Psicolog* **27**, 816-826.
- [54] Lafosse JM, Reed BR, Mungas D, Sterling SB, Wahbeh H, Jagust WJ (1997) Fluency and memory differences between ischemic vascular dementia and Alzheimer's disease. *Neuropsychology* **11**, 514-522.
- [55] Londoño D, Cifuentes V, Lubert C (2012) Correlación entre las habilidades académicas de lectura y escritura y el desempeño neuropsicológico en una muestra de niños y niñas con tdah de la ciudad de Manizales. *Psicol Caribe* **29**, 305-329.
- [56] Hills TT, Todd PM, Jones MN (2015) Foraging in semantic fields: How we search through memory. *Top Cogn Sci* **7**, 513-534.
- [57] Zemla JC (2022) Knowledge representations derived from semantic fluency data. *Front Psychol* **13**, 815860.
- [58] Collins AM, Loftus EF (1975) A spreading-activation theory of semantic processing. *Psychol Rev* **82**, 407-428.
- [59] Jescheniak JD, Levelt WJM (1994) Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *J Exp Psychol Learn Mem Cogn* **20**, 824-843.
- [60] Hurks PPM, Vles JSH, Hendriksen JGM, Kalff AC, Feron FJM, Kroes M, Zeven TMCB van, Steyaert J, Jolles J (2006) Semantic category fluency versus initial letter fluency over 60 seconds as a measure of automatic and controlled processing in healthy school-aged children. *J Clin Exp Neuropsychol* **28**, 684-695.
- [61] Klenberg L, Korkman M, Lahti-Nuutila P (2001) Differential development of attention and executive functions in 3- to 12-year-old Finnish children. *Dev Neuropsychol* **20**, 407-428.
- [62] Riva D, Nichelli F, Devoti M (2000) Developmental aspects of verbal fluency and confrontation naming in children. *Brain Lang* **71**, 267-284.
- [63] Wagner S, Sebastian A, Lieb K, Tüscher O, Tadić A (2014) A coordinate-based ALE functional MRI meta-analysis of brain activation during verbal fluency tasks in healthy control subjects. *BMC Neurosci* **15**, 19.
- [64] Schmidt CSM, Nitschke K, Bormann T, Römer P, Kümmerer D, Martin M, Umarova RM, Leonhart R, Egger K, Dressing A, Musso M, Willmes K, Weiller C, Kaller CP (2019) Dissociating frontal and temporal correlates of phonological and semantic fluency in a large sample of left hemisphere stroke patients. *Neuroimage Clin* **23**, 101840.
- [65] Meinzer M, Flaisch T, Wilser L, Eulitz C, Rockstroh B, Conway T, Gonzalez-Rothi L, Crosson B (2009) Neural signatures of semantic and phonemic fluency in young and old adults. *J Cognitive Neurosci* **21**, 2007-2018.
- [66] Kwak S, Shin SA, Ko H, Kim H, Oh DJ, Youn JH, Lee J-Y, Kim YK (2022) A comparison between the performances of verbal and nonverbal fluency tests in discriminating between mild cognitive impairments and Alzheimer's disease patients and their brain morphological correlates. *Dement Neurocogn Disord* **21**, 17-29.
- [67] Pakhomov SVS, Eberly L, Knopman D (2016) Characterizing cognitive performance in a large longitudinal study of aging with computerized semantic indices of verbal fluency. *Neuropsychologia* **89**, 42-56.
- [68] Itaguchi Y, Castro-Chavira SA, Waterloo K, Johnsen SH, Rodríguez-Aranda C (2022) Evaluation of error production in animal fluency and its relationship to frontal tracts in normal aging and mild Alzheimer's disease: A combined LDA and time-course analysis investigation. *Front Aging Neurosci* **13**, 710938.
- [69] Gladsjo JA, Schuman CC, Evans JD, Peavy GM, Miller SW, Heaton RK (1999) Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity. *Assessment* **6**, 147-178.
- [70] Aert RCM van, Wicherts JM, Assen MALM van (2019) Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis. *PLoS One* **14**, e0215052.