

<https://doi.org/10.1038/s44400-026-00113-8>

Clustering and switching patterns in semantic verbal fluency among Chinese older adults: associations with cognitive function and dementia screening



Juqiang Chen^{1,2}, Yangqing Wu^{3,4}, Xuhao Zhao^{5,6}, Xin Xu^{5,6}✉ & Mengyue Wu^{2,4}✉

This study examined semantic verbal fluency performance in Mandarin-speaking older adults from a digitally administered, community-based dementia screening project. Using *Tongyici Cilin*, a Chinese semantic knowledge base, we quantified word count, switching frequency, and cluster size in an animal fluency task. In Phase 1, a large community sample was used to assess associations between verbal fluency indices and cognitive measures of memory and orientation across age and education subgroups. Word count and switching frequency were consistently associated with cognitive performance, particularly among older adults and participants with lower education, whereas cluster size showed limited practical relevance. In Phase 2, a clinically characterized subsample was analyzed to compare participants with no cognitive impairment, mild cognitive impairment (MCI), and dementia. Both the MCI and dementia groups showed reduced word count and switching frequency relative to the no cognitive impairment group, while cluster size did not differ significantly across groups. These findings support digitally collected word count and switching frequency as useful indicators for community-based dementia screening among Chinese older adults.

Dementia represents a rapidly growing global public health challenge, with prevalence rates steadily rising among individuals over the age of 65¹. Subtle linguistic deficits, particularly in lexical access and semantic organization, often manifest years before a formal clinical diagnosis of mild cognitive impairment (MCI) or Alzheimer's disease (AD)². Consequently, language-based assessments have emerged as critical tools for early detection. Among these, the verbal fluency test, which requires individuals to generate words within a specific category (e.g., animals) under time constraints, is widely regarded as a sensitive probe of the lexical-semantic domain^{2,3}.

Despite its expediency and simplicity, the verbal fluency test involves complex cognitive processes. Individuals must access and retrieve words stored in their mental lexicon. To maximize output, participants need to switch between subcategories (e.g., moving from “pet animals” to “wild animals” in a semantic task) while inhibiting previously mentioned words to avoid repetition. The task also requires monitoring output and tracking time, thereby involving executive functions, inhibition, attention, and

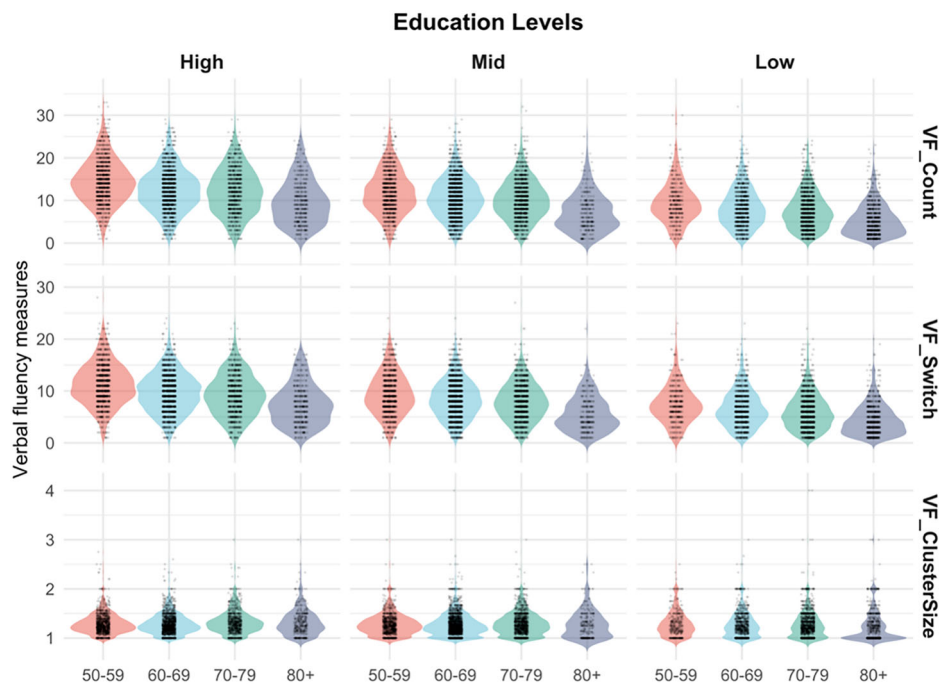
working memory^{4,5}. This task has been successfully used to examine different pathologies, including AD, Parkinson's disease^{6–9}, stroke¹⁰ and traumatic brain injury¹¹.

The most basic evaluation of verbal fluency performance is the count of correctly generated words (or word count). Different error types, such as repetitions or intrusions (e.g., words from another semantic category), can also be analyzed, particularly in populations with language or cognitive impairments. Two additional features of task performance, i.e., clustering and switching, are employed to reveal more in-depth cognitive processes underlying the task^{12–14}. Clustering refers to the production of words within subcategories (e.g., pet animals in the animal fluency task), and switching refers to the ability to shift from one subcategory to a new subcategory. It is proposed that clustering should reflect the process of accessing and using the mental lexicon, whereas switching should reflect a searching process. These two components are further hypothesized to be related to different brain regions and functions: clustering is related to temporal lobe functions, such

¹School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China. ²National Research Centre for Language and Well-being, Shanghai Jiao Tong University, Shanghai, China. ³SJTU Paris Elite Institute of Technology, Shanghai Jiao Tong University, Shanghai, China. ⁴X-LANCE Lab, School of Computer Science, MoE Key Lab of Artificial Intelligence, Shanghai Jiao Tong University, Shanghai, China. ⁵School of Public Health, The Second Affiliated Hospital of School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China. ⁶Nanhu Brain-computer Interface institute, Hangzhou, Zhejiang, China.

✉ e-mail: xuxinsummer@zju.edu.cn; mengyuewu@sjtu.edu.cn

Fig. 1 | Semantic verbal fluency performance across age groups and education levels. Violin plots show the distribution of wordcount (VF_Count), switching frequency (VF_Switch), and cluster size (VF_ClusterSize).



as verbal memory and word storage, while switching depends on frontal lobe processes, such as strategic search, cognitive flexibility, and shifting¹⁴.

In a seminal study¹³, younger participants in a semantic fluency task generated more words and switched more frequently than older participants, though no group difference was observed in cluster size. Furthermore, Fagundo et al. (2008) conducted both cross-sectional and longitudinal studies to examine verbal fluency in three groups of participants, i.e., participants with memory complaints, AD patients, and healthy controls. The cross-sectional study revealed that AD patients produced significantly fewer correct words, smaller cluster sizes, and exhibited less switching than participants with memory complaints and healthy controls. Participants with memory complaints also showed deficits in word production and clustering but not in switching compared to healthy controls. In the longitudinal follow-up, individuals from the MCG who progressed to AD exhibited a further decline in correct word count and cluster size, but not in switching performance, compared to those who did not develop dementia. The variability of clustering and switching behaviors across different ages and cognitive statuses remains incompletely understood, highlighting the need for further investigation.

Moreover, few studies have compared verbal fluency data among non-English-speaking individuals with dementia, MCI, and healthy controls. A study of Cantonese speakers¹⁵ found significant differences in verbal fluency scores between control and AD groups, as well as between mild and moderate AD subgroups. Verbal fluency scores correlated significantly with the Mini-Mental State Examination scores (the Cantonese version) for both AD and control groups. However, they did not calculate cluster size and switching times for the verbal fluency data. In another study with Mandarin Chinese participants¹⁶, they compared verbal fluency performance in terms of correct words, cluster size, subcategory (i.e., the number of different types of clusters generated) and switching times among participants with AD, vascular dementia (VaD), MCI, vascular cognitive impairment without dementia (VCIND) and cognitively normal senior controls (NC). The general pattern was that control participants performed better than participants with MCI/VCIND, who in turn performed better than participants with AD/VaD in the total verbal fluency score and all sub-scores. Sub-category and switching times were higher for AD than VaD, suggesting that these measures help distinguish between the two dementia subtypes.

So far, prior studies have demonstrated the validity of verbal fluency performance for detecting dementia and, to some extent, distinguishing its

subtypes, although most evidence comes from relatively small samples. The present study extends this line of work by analyzing animal verbal fluency data from a large, digitally collected, community-based sample of Mandarin-speaking older adults within a dementia screening project. We adopt a two-phase design. In Phase 1, we used the large community-dwelling sample to establish the empirical and cognitive validity of the verbal fluency measures. A methodological innovation of the study is the use of a comprehensive Chinese semantic knowledge base (*Tongyici Cilin*) to define animal subcategories. We examined how word count, clustering and switching indices correlated with cognitive performance (e.g., orientation and memory) across different age groups and education levels. This phase aimed to demonstrate how effectively verbal fluency measures captured meaningful variations related to cognitive function.

In Phase 2, we compared the same set of verbal fluency measures across different groups, including No Cognitive Impairment (NCI), MCI, and Dementia groups, in a clinically characterized subsample. Linear mixed-effects models were used to test group differences. By integrating large-scale community-based evidence (Phase 1) with clinical validation (Phase 2), the study demonstrates how verbal fluency features can contribute to dementia screening.

Results

This two-phase study first examined correlations between verbal fluency metrics (word count, cluster size, switching frequency) and cognitive factors (memory and orientation) in different age groups and at different education levels in a large community sample (Phase 1, $N = 10571$), then validated the discriminative utility of verbal fluency metrics in a clinically diagnosed cohort (Phase 2, $N = 311$).

Correlational analyses of the community sample

First, we used the large community-dwelling sample to establish the empirical and cognitive validity of the verbal fluency measures (see Fig. 1). Specifically, we correlated verbal fluency measures with cognitive performance measures with the aim of demonstrating how effectively verbal fluency measures capture meaningful variations related to cognitive function (Table 1).

Our results generally showed that word count and switching frequency were positively correlated with memory and orientation scores. In contrast, cluster size showed statistically significant correlations with both cognitive

measures, but these effects were very small in magnitude (small: $|r| = 0.1$, medium: $|r| = 0.3$, large: $|r| = 0.5$, Cohen, 1988)¹⁷ and may not be practically meaningful ($|r| < 0.1$). These findings support the validity of word count and switching as sensitive indicators of cognitive performance in community-dwelling older adults.

To further explore how age and education modulate these dynamics, we stratified the dataset into four age groups, each with three education levels, i.e., low (illiterate and primary school), intermediate (secondary school and high school), and high (college and above), as shown in Fig. 2.

Memory showed a significant positive correlation with word count and switching frequency, mostly with small-to-medium effect sizes. However, memory was positively correlated with cluster size only in adults aged 80 years and older with low-to-intermediate education levels.

Orientation demonstrated a significant positive correlation with word count and switching frequency across all age groups among individuals with low education. Among those with mid-level education, significant

correlations with small effect sizes ($0.1 < |r| < 0.3$) were mostly observed in participants aged 60 years and older, whereas among individuals with high education, it was observed only in those aged 80 years and older. Orientation was not significantly correlated with cluster size across most age groups and education levels. The sole exception was a very small effect ($r = 0.07$) observed in the 60–69 age group, which is unlikely to be practically meaningful.

Group differences in verbal fluency performance of the clinical cohort

We next analyzed data from the clinically diagnosed subset (Phase 2) to examine differences in verbal fluency across different cognitive status groups (see Fig. 3).

We built three separate linear mixed-effects regression (LMER) models¹⁸, each with word count, switching frequency and cluster size as the dependent variables, respectively. All models included cognitive status (NCI, MCI, dementia) and age of the subjects as a fixed factor. Genders and education levels were considered as random factors with random intercepts to maximize the models' generalizability across the data. We utilized the Kenward-Roger degrees of freedom approximation¹⁹ and the *Anova* function from the *car* package in *R*, specifying the test as "F". Main effects for multilevel factors are reported instead of comparisons with the baseline level using *t* tests. Therefore, the reported main effects were averaged across all levels of the other effects and can be directly used to test whether these verbal fluency measures can detect group differences (see Table 2). Post-hoc pairwise comparisons with Tukey's adjustment were used to identify specific group differences (see Table 3).

For word count and switching frequency, significant main effects of group and age were found, whereas the group-by-age interaction was not significant. Pairwise comparisons revealed that both the dementia and MCI groups showed lower word count and switching frequency than the NCI

Table 1 | Pearson correlation coefficients among semantic verbal fluency measures and cognitive scores in the community sample

	Switching frequency	Word count	Cluster size	Memory
Word count	0.95*			
Cluster size	-0.074*	0.18*		
Memory	0.41*	0.41*	0.047*	
Orientation	0.18*	0.18*	0.047*	0.15*

Word count: total correct animal names; switching frequency: number of transitions between semantic subcategories, cluster size: mean number of words per semantic cluster; memory: delayed recall task score (0–5); orientation: temporal/spatial orientation task score (0–6). * $p < 0.05$.

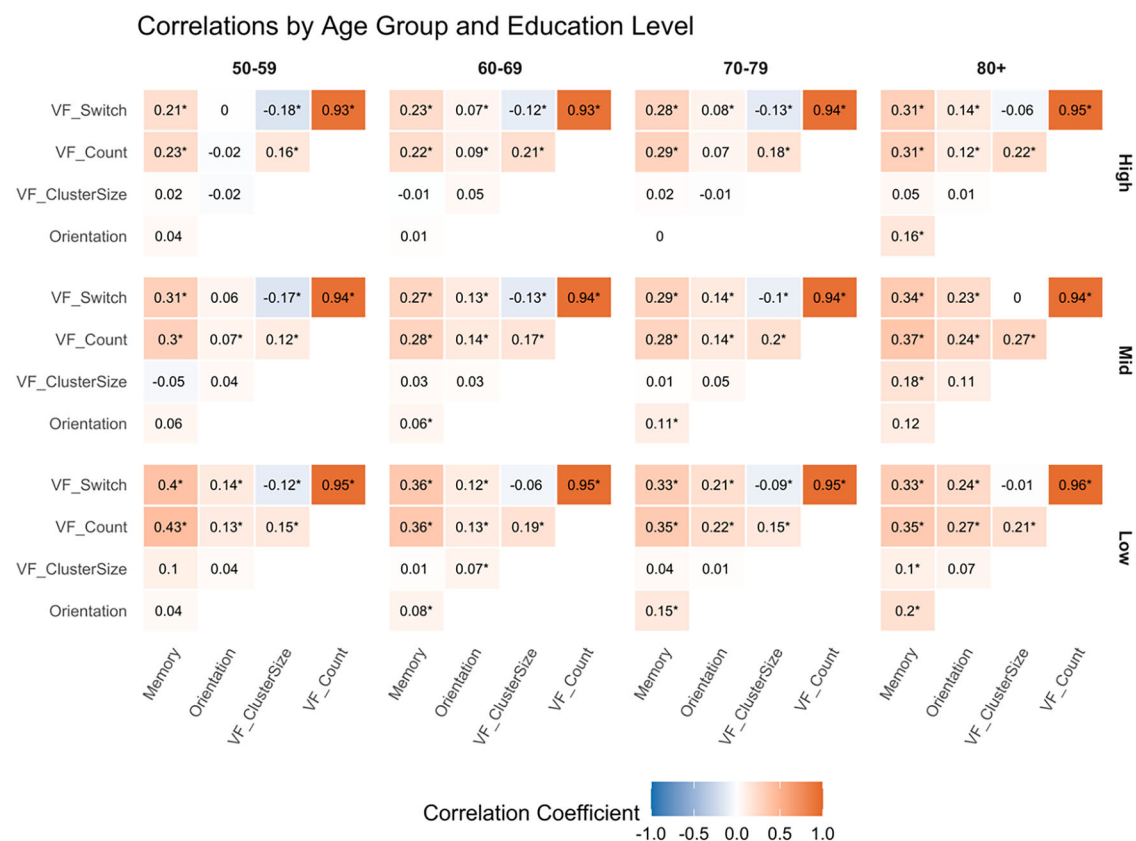


Fig. 2 | Correlations between verbal fluency performance and cognitive measures stratified by age groups and education levels. Heatmaps show the correlation patterns between verbal fluency indices and memory or orientation scores across age and education subgroups.

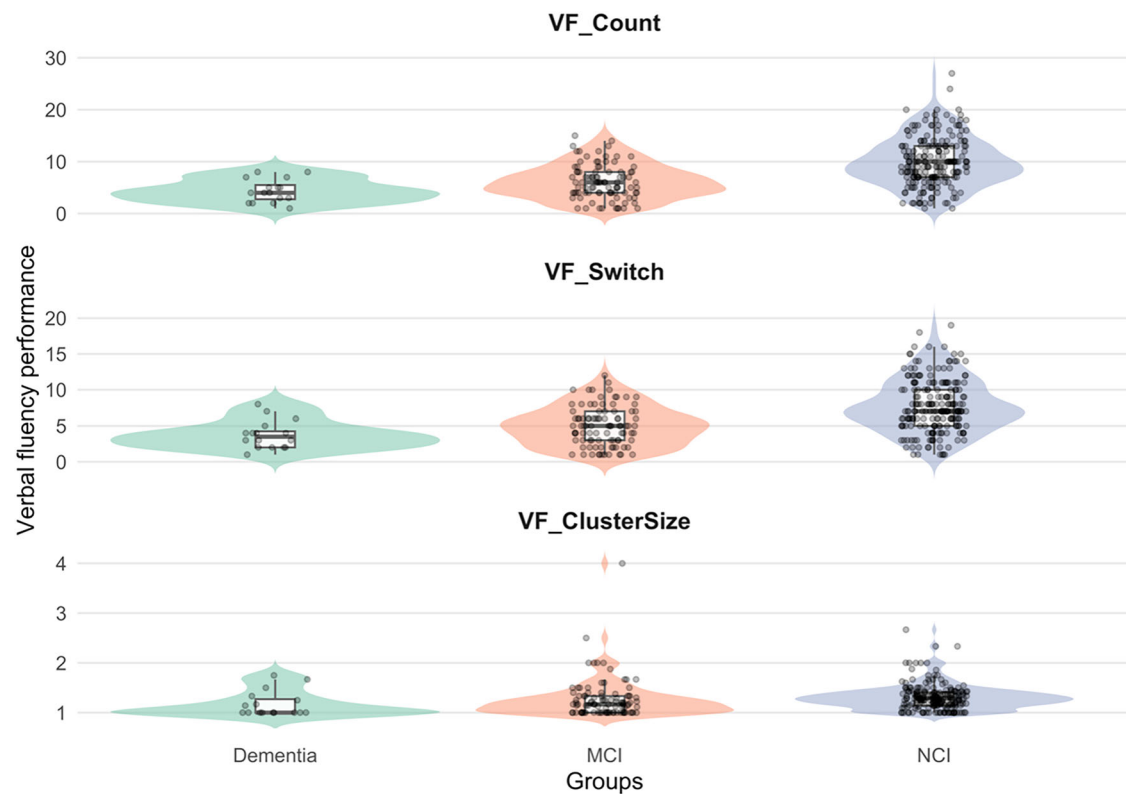


Fig. 3 | Group differences in semantic verbal fluency performance across diagnostic categories. Violin plots with overlaid boxplots showing the distribution of word count (VF_Count), switching frequency (VF_Switch), and mean cluster size (VF_ClusterSize) for the no cognitive impairment, mild cognitive impairment, and dementia groups.

Table 2 | Main effects and interactions for verbal fluency measures

Verbal fluency indices	Factors	F	df		p
Word count	Group	11.39	2	304	<0.01
	Age	5.72	1	304	0.03
	Group * Age	0.76	2	302	0.47
Switching frequency	Group	8.84	2	304	<0.01
	Age	7.59	1	304	<0.01
	Group * Age	0.24	2	302	0.79
Cluster size	Group	1.2	2	285	0.30
	Age	0.02	1	252	0.89
	Group * Age	4.74	2	303	<0.01

Significant results are in bold.

group, but no significant difference was observed between the dementia and MCI groups. Thus, word count and switching frequency appear to discriminate primarily between impaired and unimpaired individuals rather than between levels of impairment. For cluster size, no significant main effects of group or age emerged. However, a significant group-by-age interaction was detected. Pairwise comparisons between groups were uniformly nonsignificant, suggesting that the interaction reflects age-contingent effects that are not captured by group differences.

Discussion

The present study yielded findings with implications for understanding verbal fluency performance in Mandarin-speaking older adults and its utility for dementia screening. First, in a large community-dwelling sample (Phase 1), word count and switching frequency emerged as valid indicators of memory and orientation, whereas cluster size showed only limited

correlations with these cognitive factors. Second, these correlations were generally stronger among older adults and those with lower levels of education. Third, in the clinically characterized subsample (Phase 2), we observed significant group differences from normal ageing to dementia. Word count and switching frequency progressively declined from NCI to MCI and Dementia, while cluster size did not differ significantly across groups. Below, we discuss each set of findings in turn, address methodological limitations, and offer directions for future research.

Phase 1 findings established the cognitive relevance of verbal fluency measures, particularly word count and switching frequency, as valid indicators of memory and orientation in a large community-dwelling sample of older adults. Word count and switching frequency consistently captured meaningful variance in basic cognitive functions, whereas cluster size showed only negligible associations with cognitive performance. This pattern aligns with previous findings that the younger group generated more words and switched more often but displayed similar cluster size relative to the older group¹³.

In addition, when we stratified by age and education, the positive correlations between verbal fluency metrics and cognitive measures (e.g., orientation) were consistently stronger or only present among older adults and those with lower education. One cognitively relevant explanation is that older individuals with lower education may rely more heavily on basic cognitive scaffolding than younger individuals with higher education, thereby producing stronger observed correlations. Nevertheless, we acknowledge that simple orientation tasks and five-word memory tests are inherently easy for younger-old adults and those with higher education. In these subgroups, performance is likely compressed at the upper end, namely, a ceiling effect, which may attenuate correlation coefficients due to restricted range. Conversely, older adults and those with lower education exhibit fuller performance on these same measures, thereby allowing the correlations with verbal fluency to emerge more clearly.

In summary, while our findings affirm the cognitive validity of word count and switching as sensitive cognitive indicators in community-

Table 3 | Results of post-hoc pairwise comparisons (Tukey-adjusted) for verbal fluency measures across diagnostic groups

Models	Groups	Estimates	SE	df	t	p
Word count	Dementia - MCI	-1.54	1.3	302	-1.18	0.46
	Dementia - NCI	-3.75	1.28	303	-2.92	0.01
	MCI - NCI	-2.20	0.55	305	-4	<0.01
Switching frequency	Dementia - MCI	-1.10	1.03	302	-1.06	0.54
	Dementia - NCI	-2.69	1.01	303	-2.65	0.02
	MCI - NCI	-1.59	0.44	305	-3.65	0.001
Cluster size	Dementia - MCI	-0.07	0.1	303	-0.74	0.74
	Dementia - NCI	-0.09	0.1	301	-0.89	0.65
	MCI - NCI	-0.01	0.04	263	-0.31	0.95

dwelling older adults, the moderation by age and education must be interpreted with caution. Future research should employ more challenging cognitive measures to examine age- and education-related variations in verbal fluency measures.

The analysis in Phase 2 provides a deeper exploration of verbal fluency performance as a function of cognitive decline among three groups: NCI, MCI, and Dementia.

First, both the MCI and dementia groups produced significantly fewer word counts compared to the NCI group, indicating the impact of cognitive decline on verbal fluency. This result aligns with existing literature^{15,16,20–22}, suggesting a deterioration in lexical production abilities as cognitive impairment worsens. The fact that the MCI group generated significantly fewer words than the NCI group indicates that even early stages of cognitive decline can adversely affect verbal fluency. The reduction in the number of words generated across these groups suggests that verbal fluency tasks may serve as a useful tool for early detection and monitoring cognitive decline.

The analysis of clustering and switching behaviors in the verbal fluency task offers important insights into the cognitive processes affected by different stages of cognitive decline. Clustering, which refers to the production of words within subcategories, and switching, the ability to shift between these subcategories, are both critical components of verbal fluency that reveal how individuals access and organize lexical information. Our findings revealed differences in switching frequency but not cluster size between NCI and participants with MCI and dementia, generally consistent with previous studies^{16,20}. The dementia group exhibited the lowest switching frequency, indicating a reduced ability to efficiently shift between subcategories. The MCI group displayed an intermediate level of switching performance, lower than the NCI group but comparable with the dementia group.

Unlike some previous studies, our study did not find significant differences in cluster size among participants with NCI, MCI, and dementia. For instance, Fagundo et al. (2008) reported that a control group generated larger clusters than the memory complaints group (MCG) and AD group. Similarly, Gomez and White (2006) found that the control group generated larger clusters than the very mild AD group. The typical interpretation of such findings is that as cognitive decline progresses, the ability to generate semantically related words within specific subcategories deteriorates, reflecting a breakdown in the retrieval of lexical-semantic information stored in memory. Our results, however, did not replicate this pattern, suggesting that cluster size may be less sensitive to early or MCI in community-dwelling samples, or that differences in sample characteristics (e.g., group composition, severity range, or educational background) may account for the discrepancy.

It should be noted that the validity of measures, such as switching frequency and cluster size, depends on the appropriateness of defining word categories. Previous studies rely heavily on schemes that were often developed for English-speaking populations. These schemes may not fully capture the semantic relationships in other languages, potentially leading to inaccuracies when applied cross-culturally. Our use of a semantic knowledge base, such as *Cilin* provides a relatively objective classification of words

in our automated pipeline for generating verbal fluency measures. Future studies could further explore this approach in other languages and also examine other computational ways of representing lexical knowledge^{23,24}.

However, several limitations should be considered. One limitation of this study is the imbalance in group sizes, particularly the very small Dementia group compared to the NCI and MCI groups. While the observed pattern, with the Dementia group performing worst, the MCI group intermediate, and the NCI group best on measures such as word count and switching frequency, is clinically plausible, the small Dementia sample size raises concerns regarding the stability of the estimates for that group and the overall robustness of the group comparisons. Encouragingly, the current findings provide solid proof of concept and a strong foundation for better-powered replication efforts with larger, more balanced cohorts, which will further validate and refine these promising group differences.

Another limitation of this study is the substantial gender imbalance within the sample, with females outnumbering males. To partially account for this imbalance, we included gender as a random intercept in our statistical models, which helps control for variance attributable to gender at the individual level. However, this analytical adjustment does not fully resolve the concerns regarding generalizability. Future studies should aim for more balanced sex representation or conduct sex-stratified analyses to confirm whether the observed effects hold equally across genders.

A third limitation concerns the use of automatic speech recognition (ASR) for transcribing the animal naming test in Mandarin. Our empirical validation of a subset of recordings yielded an estimated accuracy of 83.2% for the ASR model used (Alibaba Cloud), which is lower than the accuracy (98%) reported under controlled laboratory conditions. ASR errors, such as misrecognition of valid animal names as non-words or omission of responses, may introduce noise into verbal fluency indices. However, this limitation is largely unavoidable in large-scale, community-based dementia screening, where manual transcription of thousands of verbal fluency responses is impractical due to time and cost constraints. It is important to note that ASR technology is advancing rapidly, with ongoing improvements in handling background noise, speaker accents, and aging voices. As ASR systems continue to evolve, particularly with the development of domain-specific models trained on older adult speech and clinical populations²⁵, measurement error will likely diminish substantially.

In conclusion, among Mandarin-speaking older adults, particularly those with lower educational attainment, verbal fluency measures (word count and switching frequency) correlate significantly with cognitive functions such as memory and orientation. Individuals with MCI and dementia exhibited marked reductions in word count and switching frequency, reflecting underlying impairments in lexical organization and cognitive flexibility. Notably, the presence of such deficits already in the MCI group suggests that disruptions in word generation and switching emerge early in the trajectory of cognitive decline. These findings support the use of digitally administered verbal fluency assessments for large-scale dementia screening in community populations.

Table 4 | Demographic characteristics of the community-dwelling older adult sample, stratified by age group and education level (Phase 1)

	50–59 (N = 2183)	60–69 (N = 3977)	70–79 (N = 3139)	80+ (N = 1272)	p ¹
Gender					<0.001
Female	1541 (71%) ²	2472 (62%)	1898 (60%)	719 (57%)	
Male	642 (29%)	1505 (38%)	1241 (40%)	553 (43%)	
Age	55(3)	65 (3)	74 (3)	84 (3)	<0.001
Education level					<0.001
High	946 (43%)	1315 (33%)	762 (24%)	365 (29%)	
Mid	891 (41%)	1812 (46%)	1132 (36%)	248 (19%)	
Low	346 (16%)	850 (21%)	1245 (40%)	659 (52%)	

¹n (%); mean (SD).²Pearson's Chi-squared test; one-way ANOVA.**Table 5 | Demographic and clinical characteristics of the participants in the diagnostic validation phase (Phase 2, N = 311)**

	Dementia (N = 16) ¹	MCI (N = 92)	NCI (N = 203)	p ²
Gender				0.15
Female	10 (62%) ¹	58 (63%)	149 (73%)	
Male	6 (38%)	34 (37%)	54 (27%)	
Age	78 (10)	74 (9)	69 (7)	<0.001
Education				<0.001
High	1 (6.3%)	5 (5.4%)	47 (23%)	
Mid	2 (13%)	27 (29%)	96 (47%)	
Low	13 (81%)	60 (65%)	60 (30%)	

NCI no cognitive impairment, MCI mild cognitive impairment, and dementia groups.

¹n (%); mean (SD).²Fisher's exact test; one-way ANOVA.**Table 6 | Hierarchical structure of Tongyici Cilin (Cilin) with an example code**

Level	1	2	3	4	5
Category	Major class	Intermediate class	Minor class	Group	Atomic group
Code	B	i	0 6	A	0 1

The five-tier structure (major class, intermediate class, minor class, group, atomic group) was used to objectively define semantic categories for animal names in the verbal fluency task.

Methods

Data collection

The data we used was obtained via a Digital Cognitive Screener (or DCS)²⁶, which included three tasks. The first task was a delayed memory test in which participants had to recall 5 words (i.e., face, silk, chrysanthemum, hotel, red) with a total score of 5. The second was an orientation task in which participants needed to answer six questions about the current year, month, week, date, city and address, with a total score of 6. The third is a semantic verbal fluency task in which sixty seconds were allotted for participants to generate animal names. This study received approval from the Medical Ethics Committee at Zhejiang University School of Public Health (ZGL202101-1), and written informed consent was obtained from all participants or their legally authorized representatives.

The DCS platform employs voice-recognition technology and is designed for utilization by non-professional facilitators, including caregivers or social workers²⁶. The platform autonomously evaluates participants' performance by first converting the test audio into text through automatic speech recognition (ASR). The system also includes a named entity recognition model to identify crucial details, such as year, day, and address. The extracted information is cross-referenced with the accurate standard answer. For every question, the system autonomously assesses the response and assigns a corresponding score. A previous study has confirmed the validity of this system in terms of the consistency between automated scoring and the conventional manual approach²⁶.

Participants

In Phase 1, we selected participants between the ages of 50 and 100 from the dementia screening project and excluded those with missing data. We then regrouped the dataset (N = 10571) into four age groups (see Table 4). Education level was originally coded as: 1 = illiterate, 2 = primary school, 3 = secondary school, 4 = high school, 5 = college and above. We regrouped these into three education levels: low (illiterate and primary school), intermediate (secondary school and high school), and high (college and above).

In Phase 2, a subset of participants' data (N = 311) from Phase 1 received comprehensive cognitive evaluations, including Montreal Cognitive Assessment (MoCA)²⁷, the Clinical Dementia Rating (CDR) scale²⁸, and a comprehensive neuropsychological evaluation, from licensed professionals under double-blind conditions, and then a unified dementia classification was reached through a joint clinical consensus process. The inclusion criterion was the provision of informed consent with willingness to participate. Exclusion criteria were: (1) severe visual, auditory, or communication impairments; (2) major depressive disorder or other psychiatric illnesses; and (3) refusal to participate. These participants were then classified into three groups: NCI, MCI, or dementia. Notably, due to the absence of biomarker confirmation, we were unable to restrict the dementia group to AD alone nor to distinguish mixed etiologies. Therefore, the broader clinical label Dementia is used throughout the results section. Table 5 summarizes the demographic characteristics of the clinical cohort.

Calculating clusters and switches

Many tools²⁹ have been developed to facilitate the analysis of verbal fluency data. Scheme files are required to compute cluster sizes and switches in these tools. There exist several English animal category schemes from previous studies^{13,30}. However, these schemes rely on subjective judgements and may not be valid across different languages and cultures.

We employed *Tongyici Cilin* (or *Cilin* in short), an existing comprehensive semantic knowledge base, as an objective and comprehensive way of defining semantic categories. *Cilin* is a dictionary created by HIT IR-Lab and organized in a five-tier hierarchical tree structure (see Table 6). The first tier indicates major classes, and the second tier is the intermediate level, e.g., Bh for all plants and Bi for all animals. The third and fourth tiers further narrow down the category. For example, horse is marked with Bi06A whereas donkey is marked with Bi06B. The fifth tier further distinguishes subcategories, e.g., foals (= Bi06A02) and mares (= Bi06A03). The words grouped at the terminal nodes have the same or similar meaning, e.g., “驴子”, “毛驴”, “驴” (= Bi06B01, donkey in Chinese). This dictionary was originally created for and has been widely used in Chinese natural language processing³¹.

The dataset was inspected for unregistered, ambiguous, and polysemous lexical items. Lexical items not represented in the *Cilin* taxonomy were assigned to their semantically nearest hypernym or superordinate category. For entries containing modifiers (e.g., 小猫, “little cat”), the modifier was removed to retain the base animal term (e.g., 猫, “cat”). In cases where a word carried both metaphorical and denotative animal meanings (e.g., “鸳鸯”, yuān yāng, which metaphorically refers to a romantic couple and denotatively refers to the mandarin duck), the literal

animal interpretation was retained. Repeated responses were not excluded from the analysis.

Clusters were defined as groups of successively generated words belonging to the same semantic subcategory with the same first three levels based on *Cilin*. For example, a horse and a donkey were treated as belonging to the same category. Cluster size was counted from the first word in each cluster, and the mean cluster size was calculated for each participant. Switches were calculated as the number of transitions between clusters, including single words.

Statistical analyses

In Phase 1, Pearson correlation analyses were conducted between verbal fluency measures (word count, switching frequency, cluster size) and cognitive measures (memory and orientation), initially across the entire dataset and subsequently within age- and education-stratified subsamples. This stratified approach was adopted to examine how the observed correlations varied across different age and education subgroups.

In Phase 2, three separate LMER models were constructed, with word count, switching frequency, and cluster size entered as dependent variables, respectively. Each model included cognitive status (NCI, MCI, dementia) and participant age as fixed factors. Gender and education level were treated as random factors with random intercepts to enhance the generalizability of the models across the dataset. The Kenward-Roger degrees of freedom approximation¹⁹ was applied, and the *Anova* function from the *car* package in R was used with the test specified as “F”. Main effects for multilevel factors are reported rather than *t* test-based comparisons against a baseline level. Consequently, the reported main effects represent averages across all levels of other effects and can be directly interpreted to test whether the verbal fluency measures are sensitive to group differences. Post-hoc pairwise comparisons with Tukey’s adjustment were conducted to identify specific group differences.

Data availability

The data that support the findings of this study are available from the authors, Xin Xu and Mengyue Wu, upon reasonable request.

Code availability

The code used in this study is available at OSF: https://osf.io/bx4s3/overview?view_only=9a6b8a418fcd4237bb09b0645f03a602.

Received: 14 February 2026; Accepted: 5 June 2026;

Published online: 08 July 2026

References

- Jia, L. et al. Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: a cross-sectional study. *Lancet Public Health* **5**, e661–e671 (2020).
- Amieva, H. et al. Prodromal Alzheimer’s disease: successive emergence of the clinical symptoms. *Ann. Neurol.* **64**, 492–498 (2008).
- Salehi, M., Reisi, M. & Ghasisin, L. Lexical retrieval or semantic knowledge which one causes naming errors in patients with mild and moderate Alzheimer’s disease. *Dement. Geriatr. Cogn. Disord. Extra* **7**, 419–429 (2017).
- Aita, S. L. et al. Executive, language, or both? An examination of the construct validity of verbal fluency measures. *Appl. Neuropsychol. Adult* **26**, 441–451 (2019).
- Laisney, M. et al. The underlying mechanisms of verbal fluency deficit in frontotemporal dementia and semantic dementia. *J. Neurol.* **256**, 1083–1094 (2009).
- Bertola, L. et al. Graph analysis of verbal fluency test discriminate between patients with Alzheimer’s disease, mild cognitive impairment and normal elderly controls. *Front. Aging Neurosci.* **6**, 185 (2014).
- Fagundo, A. B. et al. Clustering and switching in semantic fluency: predictors of the development of Alzheimer’s disease. *Int. J. Geriatr. Psychiatry* **23**, 1007–1013 (2008).
- Henry, J. D., Crawford, J. R. & Phillips, L. H. Verbal fluency performance in dementia of the Alzheimer’s type: a meta-analysis. *Neuropsychologia* **42**, 1212–1222 (2004).
- Monsch, A. U. et al. Comparisons of verbal fluency tasks in the detection of dementia of the Alzheimer type. *Arch. Neurol.* **49**, 1253–1258 (1992).
- Schmidt, C. S. M. et al. Dissociating frontal and temporal correlates of phonological and semantic fluency in a large sample of left hemisphere stroke patients. *NeuroImage Clin.* **23**, 101840 (2019).
- Henry, J. D. & Crawford, J. R. A meta-analytic review of verbal fluency performance in patients with traumatic brain injury. *Neuropsychology* **18**, 621–628 (2004).
- Troyer, A. K. Normative data for clustering and switching on verbal fluency tasks. *J. Clin. Exp. Neuropsychol.* **22**, 370–378 (2000).
- Troyer, A. K., Moscovitch, M. & Winocur, G. Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology* **11**, 138–146 (1997).
- Troyer, A. K., Moscovitch, M., Winocur, G., Alexander, M. P. & Stuss, D. Clustering and switching on verbal fluency: the effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia* **36**, 499–504 (1998).
- Mok, E. H. L., Lam, L. C. W. & Chiu, H. F. K. Category verbal fluency test performance in Chinese elderly with Alzheimer’s disease. *Dement. Geriatr. Cogn. Disord.* **18**, 120–124 (2004).
- Zhao, Q., Guo, Q. & Hong, Z. Clustering and switching during a semantic verbal fluency test contribute to differential diagnosis of cognitive impairment. *Neurosci. Bull.* **29**, 75–82 (2013).
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. (Routledge, 1988).
- Bates, D. M. *lme4: mixed-effects modeling with R* (2015).
- Halekoh, U. & Højsgaard, S. A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models - the R Package pbkrtest. *J. Stat. Softw.* **59**, 1–32 (2014).
- Epker, M. O., Lacritz, L. H. & Munro Cullum, C. Comparative analysis of qualitative verbal fluency performance in normal elderly and demented populations. *J. Clin. Exp. Neuropsychol.* **21**, 425–434 (1999).
- Farzanfar, D., Statucka, M. & Cohn, M. Automated indices of clustering and switching of semantic verbal fluency in Parkinson’s disease. *J. Int. Neuropsychol. Soc.* **24**, 1047–1056 (2018).
- Gomez, R. & White, D. Using verbal fluency to detect very mild dementia of the Alzheimer type. *Arch. Clin. Neuropsychol.* **21**, 771–775 (2006).
- Alacam, Ö., Schüz, S., Wegrzyn, M., Kißler, J. & Zarriß, S. Exploring Semantic Spaces for Detecting Clustering and Switching in Verbal Fluency. in *Proceedings of the 29th International Conference on Computational Linguistics* (eds. Calzolari, N. et al.) 178–191 (International Committee on Computational Linguistics, 2022).
- Kim, N., Kim, J.-H., Wolters, M. K., MacPherson, S. E. & Park, J. C. Automatic scoring of semantic fluency. *Front. Psychol.* **10**, 1020 (2019).
- Gohider, N. & Basir, O. Benchmarking Self-Supervised Speech Models on the Disordered Speech Recognition Task. in *2026 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)* 1–8 (2026).
- Zhao, X. et al. A voice recognition-based digital cognitive screener for dementia detection in the community: Development and validation study. *Front. Psychiatry* **13**, 899729 (2022).
- Nasreddine, Z. S. et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **53**, 695–699 (2005).

28. Morris, J. C. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* **43**, 2412–2414 (1993).
29. Zemla, J. C., Cao, K., Mueller, K. D. & Austerweil, J. L. SNAFU: the semantic network and fluency utility. *Behav. Res.* **52**, 1681–1699 (2020).
30. Hills, T. T., Jones, M. N. & Todd, P. M. Optimal foraging in semantic memory. *Psychol. Rev.* **119**, 431–440 (2012).
31. Wang, X., Wang, X. & Liu, Z. Cilin-Based Semantic Similarity Calculation of Chinese Words. in *Advances in Intelligent Systems and Computing* (eds. Qiao, F., Patnaik, S. & Wang, J.) 240–244 (Springer International Publishing, 2018).

Acknowledgements

This work was supported by Key R&D Program of Zhejiang (2025C02108), Guangxi major science and technology project (No. AA23062062), the China NSFC Project (No. 62572320), the Humanities and Social Sciences Foundation of the Ministry of Education, PRC (23YJC740002) and the National Language Commission (YB145-166). The authors thank all participants and community workers involved in the dementia screening project.

Author contributions

J.C. and M.W. conceptualized the work. X.X. and X.Z. were responsible for data collection. J.C. and Y.W. developed the code and conducted the computational analysis. Y.W. performed the machine learning modeling. J.C. wrote the first draft of the manuscript, and all authors participated in the revision of the manuscript. All authors read and approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Xin Xu or Mengyue Wu.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026