Comprehensive verbal fluency features predict executive function performance

Julia Amunts*[1,2], Julia A. Camilleri[1,2], Simon B. Eickhoff[1,2], Kaustubh R. Patil[1,2], Stefan Heim[3,4], Georg G. von Polier[1,5,6] & Susanne Weis[1,2]

[1] Institute of Neuroscience and Medicine (INM-7 Brain and Behaviour), Forschungszentrum Jülich, Germany
[2] Institute of Systems Neuroscience, Heinrich-Heine University, Düsseldorf, Germany
[3] Institute of Neuroscience and Medicine (INM-1 Structural and functional organization of the brain), Forschungszentrum Jülich, Germany
[4] Department of Psychiatry, Psychotherapy und Psychosomatics, Medical Faculty, RWTH Aachen University, Aachen, Germany
[5] Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Goethe-Universität Frankfurt am Main, Frankfurt am Main, Germany
[6] Department of Child and Adolescent Psychiatry, Psychosomatic and Psychotherapy, University Hospital RWTH Aachen, Germany

Corresponding author:
Julia Amunts
Institute of Neuroscience and Medicine (INM-7 Brain and Behaviour), Research Center Jülich, Jülich, Germany
Wilhelm-Johnen-Str.
52428 Jülich
GERMANY

Correspondence to j.amunts@fz-juelich.de

Abstract

Semantic verbal fluency (sVF) tasks are commonly used in clinical diagnostic batteries as well as in a research context. When performing sVF tasks to assess executive functions (EFs) the sum of correctly produced words is the main measure. Although previous research indicates potentially better insights into EF performance by the use of finer grained sVF information, this has not yet been objectively evaluated.

To investigate the potential of employing a finer grained sVF feature set to predict EF performance, healthy monolingual German speaking participants (n=230) were tested with a comprehensive EF test battery and sVF tasks, from which features including sum scores, error types, speech breaks and semantic relatedness were extracted. A machine learning method was applied to predict EF scores from sVF features in previously unseen subjects. To investigate the predictive power of the advanced sVF feature set, we compared it to the commonly used sum score analysis.

Results revealed that 8 / 14 EF tests were predicted significantly using the comprehensive sVF feature set, which outperformed sum scores particularly in predicting cognitive flexibility and inhibitory processes.

These findings highlight the predictive potential of a comprehensive evaluation of sVF tasks which might be used as diagnostic screening of EFs.

Introduction

Executive functions (EFs) comprise cognitive processes that enable goal directed behaviour [1]. Previous literature investigated the general cognitive processes that fall under the umbrella term of EFs and encompass both lower-level cognitive processes and higher-level processes. The former include working memory, inhibition and cognitive flexibility which represent the building blocks for higher-level processes such as planning, reasoning and problem solving [2]. While the number and definition of different EF subprocesses remains controversial [3], there is strong evidence that EFs are impaired in a large number of neurological [4,5] and psychiatric [6,7] diseases. Therefore, the measurement of EFs forms a crucial part of the clinical neuropsychological diagnostical routine in order to detect and specify impairments such as frontal lobe damages [8]. Multiple test batteries such as the *Delis-Kaplan Executive Function System* (D-KEFS) [9] and the *Vienna Test System* [10] provide numerous EF tests to capture a wide range of the different aspects of EFs. However, many EF tests are mainly based on pen-and-paper versions which tend to be time consuming while also lacking accuracy. Moreover, there are discrepancies between unnatural test instructions and naturalistic tasks in everyday life which leads to a lack of ecological validity of commonly used EF tests [11].

There is consensus, that EFs play a crucial role in speech production processes [12,13]. Cognitive flexibility is required to activate general lexical concepts while later working memory capacities are needed for remembering already produced words. Here, the episodic buffer and phonological loop, which are also related to the working memory system, serve as central components [12]. Since EFs are also involved in speech production, verbal fluency (VF) tests are integrated in several clinical diagnostic batteries to assess EFs. E.g. in *B-CATS* – an assessment tool for schizophrenia; *NIH stroke scale* – assessment for quantifying stroke severity; BCSB – screening for mild Alzheimer´s disease; and D-KEFS – broadly applicable assessment used for assessing diseases such as epilepsy or Parkinson´s disease.

Two different types of VF tests are commonly used. On the one hand, VF is assessed with a lexical task. In this task participants are asked to produce as many words as possible with a specific initial letter within a specific timeframe (usually 1-2 minutes). Due to the fact that all requested words start with the same phoneme, the lexical VF task is also commonly referred to as *phonological* VF task. On the other hand, the semantic VF (sVF) task requires the production of words belonging to a specific category (e.g. animals or fruits), regardless of the initial letter of the word. The lexical VF task is driven by phonological and lexical cues, whereas the sVF task requires attributes of a specific semantic category.

Within each type of the VF task, it is also possible to modulate the demand on EFs by applying a switching component. Here, participants are asked to switch between two different categories in alternating order within the same task (e.g. fruits-jobs). VF performance is generally evaluated by calculating the total number of correctly produced items. However, in the neurological literature, it has been shown that specific patterns of VF impairment greatly depend on the damaged brain regions [14,15]. Thus, studies suggest the need for a more differentiated analysis of VF performance [16].

In general, there is consensus on the involvement of EFs in the VF task in healthy controls [17] as well as their impairment in patients [15,18]. In detail, it is assumed that semantic knowledge and memory as well as cognitive flexibility are required to build semantic associations in sVF tasks whereas the lexical VF tasks require the suppression of grouping words with shared associations [19]. Additionally, in both types of VF tasks, inhibition is presumably needed to suppress competitive responses and to avoid perseveration errors [20,21], while attention, updating and working memory processes are simultaneously involved to keep the processing speed high, to remember already produced items and to produce as many items as possible [12].

Although previous findings undergirded the involvement of executive control processes in the VF task [22], the diagnostic validity of VF tasks to assess EF performance remains controversial

[23,24]. In particular, it has been found to be affected by multiple factors such as the underlying language component in the VF task, underlying cognitive processes such as intelligence, and fluctuating hormonal levels [25,26]. Moreover, the literature is not in agreement with regards to the specific relationship between VF and EF. Various studies report a positive correlation between working memory, inhibition, cognitive flexibility performance and the total score of produced words [22,27]. In contrast, other studies failed to identify a clear relationship between VF performance and EFs in one or more EF domains [24,28]. Notably, in previous studies, classical statistic methods were used to e.g. investigate group comparisons of EF performance in patients and healthy controls. Applying correlational analyses, studies investigated linear relationships of VF sum scores and different EF domains [29].

However, within the last years VF tasks per se have gained more interest as a predictive tool for clinical decision making, e.g. in schizophrenia [30] or mild cognitive impairments [31] since they offer an alternative to the highly time-consuming testing procedure of EFs [11]. The growing interest in the predictive value of VF tasks might be a result of the increasing use of machine learning algorithms investigating speech production to predict disease specific properties [32–34]. The main appeal of the machine learning approach is its ability to train a predictive model by identifying patterns in high dimensional data which can be subsequently used to make predictions in unseen data. Additionally, interpreting models can provide information with regards to which specific features contribute most to accurate predictions. Based on a data-driven learning, predictive modelling enables researchers to capture (non)-linear relationships, generalize associations and to potentially subsequently transfer these to a clinical context.

Although the VF task is commonly evaluated based on the total sum of correct produced words [35–37], other variables can also be employed to gain deeper insights into cognitive performance.

Recent studies have demonstrated the potential of advanced parameters taken from the VF task, i.e. error types [38], latencies [39] and semantic distances [18] to complement the common analysis of the total sum of words. These additional variables, assessed within the sVF task, were shown to reflect the complex involvement of executive processes in disorders such as dementia as well as in better differentiation between patients and healthy controls [16,40].

To interpret VF performance in more detail, studies have also investigated error types that occurred in the course of the VF task such as those based on the breaking of sVF-specific rules (e.g. naming words from a different category, creating neologisms) and category errors. Perseveration and category errors are particularly reported in the switching VF task when participants fail to switch to the second category, name words from a different category or repeat the same category twice [37]. Thus, perseveration and category errors can provide qualitative information when measuring VF performance, in addition to the commonly used total sum of words.

Additionally, information of the VF task can also be assessed on a semantic level, analysing semantically relatedness of produced words. This concept was first investigated by Troyer et al. [41] who manually organized produced words in the sVF task into conceptually related clusters and switches. Specifically, semantic related words were clustered based on specific subcategories [41]. For example, animals were clustered based on their living environment, human use and zoological categories. According to these clusters, which are usually defined as a minimum of a two-word-sequence within the same subcategory, switches were calculated as the total number of shifts between these clusters [41]. Here, two types of switches were defined: While *cluster switches* describe a transition between multiword and adjacent clusters, *hard switches* represent transitions between a cluster and non-clustered words [42]. Later research showed that the ability to create new subcategories and generate new cues is more important for performing the sVF task than creating large cluster sizes [43]. Moreover, authors highlighted

the importance of working memory capacity for self-generating category cues in healthy participants [43] and suggested the sVF task as a diagnostic tool in cognitive impairment [44,45]. Nevertheless, this assessment of semantic information from the sVF task was traditionally done manually and thus was highly time-consuming and partially subjective due to the manual determination and assignment of subcategories [46,47]. However, this problem can be addressed with the help of computational linguistics providing automated computational approaches (e.g. Latent semantic analysis [48], Word2Vec [49]). Nowadays large text corpora and fine grained information of semantic relatedness are available (e.g. WordNet [50], DISCO [51]). In general, different conceptual structures are implemented in these models. On the one hand, some systems provide the hierarchical structure of a lexical semantic net [52] based on semantic concepts (e.g. fishes, birds, mammals) [50]. In contrast to this hierarchical and ontological approach, vector-based systems rely on the co-occurrence of words within a big text corpus. Here, words are represented as a point in a multi-dimensional space creating word embeddings [53]. Applying these computerized and automated systems, studies were able to identify dementia risk in healthy participants based on semantic relatedness [16] and to distinguish between patients with forms of disorganization and healthy controls [54].

Alongside the semantic information, the sVF task also provides prosodic information such as speech latencies (speech pauses between each word). Latencies convey information about the approximate time needed to access lexical items [13,55]. Although there is little literature on the relationship between speech latencies in the VF task and EF performance, some findings indicate that it might be meaningful [39,43]. Specifically, studies suggest that a higher incidence of unfilled pauses are more likely to occur in situations in which participants are confronted with a higher planning load [56]. Other studies also report a relationship between prosodic information and EF demands showing a decreased production of words within the progress of the VF task [39]. Since a decrease of the number of produced words in the VF task also indicates

an increase of speech latencies [39] these findings suggest that speech latencies could provide additional information on VF performance with respect to the involvement of EFs.

In summary, previous studies indicate the potential of additional quantitative measures for evaluating sVF performance to gain better insight into cognitive processes. However, diagnostic batteries used in the clinical context as well as in the scientific environment still heavily rely on the sole use of the sum of correct words as the main indicator of EF performance. Consequently, the aim of the present study was to investigate the predictive power of a comprehensive set of sVF measures and compare it to the commonly used sum score analysis. As a first step into deeper insights of the predictive power of the VF task, we focus on the semantic VF task which allowed us to exploit the vast information within the semantic relatedness features. In this exploratory study, machine learning methods were applied to predict performances of well validated but highly time-consuming EF tests from a broad set of objective and mainly computerized VF measurements in unseen participants. We expected the extended sVF feature set to outperform the basic analysis of sum scores in predicting EF test results.

Methods

Participants

In this study, 230 healthy participants with an age range of 20-55 years (mean age 35.2 ± 11.1; 92 males) were tested. Before the actual testing session, participants were asked for previously detected diagnoses. Only participants without neurological or psychiatric diagnoses were included in this study. Moreover, participants were monolingual German speakers, i.e. their native language was German and they did not learn an additional language before going to school. Participants received different levels of education (finished middle school: 8, professional school/job training: 63, finished high school with a university-entrance diploma: 69, university degree: 90). The recruitment took place in North Rhine-Westphalia (Germany) via social networks and the Forschungszentrum Jülich mailing list. Participants were tested at the Forschungszentrum Jülich, and the testing session included an EF test battery together with VF tasks, with a duration of 150-180 minutes depending on the individual time needed for instructions and the speed with which the participants passed the tests. A remuneration fee of €50 was paid. All experiments were performed in accordance with relevant guidelines and regulations. Moreover, informed consent was obtained from all participants. Collection and analyses of the data presented here was approved by the ethics committee at Heinrich-Heine University Düsseldorf.

Executive function assessment

The EF test battery consisted of 14 computerized versions of commonly used neuropsychological tests covering domains of cognitive flexibility, working memory and inhibition. While 11 of these tests were taken from the *Vienna Testsystem* [57] , three were designed with *PsyToolkit* [58]. The *Vienna Testsystem* [57] is a standardized computerized test battery providing numerous EF tests and test manuals. Every EF test provided multiple variables which were extracted automatically by the respective test system. While some of these

variables represent main variables, others solely include processing time information which are not directly linked to the EF performance. EF tests which were designed within *PsyToolkit* [58] do not come with associated test manuals and the selection of variables of these tests was thus based on previous literature [59–61].

Cognitive flexibility was assessed using five tests, namely, the *Trail Making Test* [62] (TMT), *Raven´s Standard Progressive Matrices* [63] (SPM) , *Wisconsin Card Sorting Test* [64] (WCST), *Tower of London* (TOL) [65] and *Cued-Task Switching* [66] (SWITCH).

Working memory performance was examined using three tests: *N-back non-verbal Test* [67] (NBN), *Non-verbal Learning Test* [68] (NVLT) and *Corsi Block Tapping Test* [69] (CORSI).

Inhibition was tested using *Stop-Signal Task* [70] (STOP), *Simon Task* [71] (SIMON) and *Stroop Test* [72] (STROOP).

Additionally, we also assessed divided and spatial attention (WAF-G [73], WAF-R [73]) as well as vigilance (*Mackworth Clocktest* [61]) (CLOCK). In total, 68 variables were extracted from EF tests. The full set of EF test variables is provided in the supplementary material (Table S1).

Semantic verbal fluency tasks

The sVF tasks were based on the *Regensburger Wortflüssigkeitstest* [74] (RWT) which is equivalent to the English *Controlled Oral Word Association Test* [75] (COWAT). The German standardized neuropsychological version of the VF task was used due to language-specific differences in the frequency and usage of letters and categories [76]. Two of the tasks were simple sVF tasks in which the participant had to name animals (t_1) and jobs (t_2). The third sVF task (t_3) was a switching task in which the participant switched between fruits and sports within the same task. Each of the three tasks was performed for 2 minutes. The sVF tasks were presented with *Presentation* software [77] and the participant´s responses were recorded automatically.

Following the testing session, the recorded speech was transcribed and words were coded manually as being either **correct answers** or **errors**. Furthermore, errors were differentiated

into perseveration and category errors. Sum scores of each sVF tasks separately, sum score of correct produced words across all sVF and errors (perseveration, category errors) were included in the prediction analysis. In general, the sum scores solely include correct produced items in all three sVF tasks. A list of extracted sVF features is shown in Table 1.

Table 1: Overview of Verbal fluency features

VF features	Description
Correct words t1+t2+t3	Sum of all correct produced words in task1, task2, task3
Correct words	Sum of correct produced words in each task
Switch coefficient	Relationship of correct items in simple and switching tasks; switching coefficient = sum3/((sum1+sum2)/2))
Repetition error	Repetition errors in task 1, task 2
Category error	Category errors in task 3
Latency mean	Mean of speech breaks in each task
Latencies 1st quarter	Mean of speech breaks in seconds 0-30 (i1) for each task
Latencies 2nd quarter	Mean of speech breaks in seconds 31-60 (i2) for each task
Latencies 3rd quarter	Mean of speech breaks in seconds 61-90 (i3) for each task
Latencies 4th quarter	Mean of speech breaks in seconds 91-120 (i4) for each task
Latency difference	Progress of speech breaks (i4-i1) in each task
Sequential mean	Semantic mean of all sequential word pairs in each task; computed with GermaNet (hierachical)
Cumulative mean	Semantic mean of all possible word pairs (cumulative) in each task; computed with GermaNet (hierarchical)
Sequential mean cat1 t3	Semantic mean of all sequential word pairs (sequential) in catergory 1 (sports) of switching task; computed with GermaNet (hierarchical)
Sequential mean cat2 t3	Semantic mean of all sequential word pairs (sequential) in catergory 2 (fruits) of switching task; computed with GermaNet (hierarchical)
Sequential mean DIS	Semantic mean of all sequential word pairs in each task; computed with DISCO (Word2Vec)
Cumulative mean DIS	Cumulative mean of all possible word pairs in each task; computed with DISCO (Word2Vec)
Sequential mean cat1 t3 DIS	Semantic mean of all sequential word pairs in category 1 (sports) of switching task; computed with DISCO (Word2Vec)
Sequential mean cat2 t3 DIS	Semantic mean of all sequential word pairs in category 2 (fruits) of switching task; computed with DISCO (Word2Vec)

Speech latencies were automatically detected and manually corrected using *PRAAT* [78], and the mean of the speech **latencies** within each task was calculated. Moreover, the task was divided into four 30-seconds intervals (i_1, i_2, i_3, i_4) and the mean of the speech latencies within each interval was determined. Additionally, these means of intervals were then used to determine an increase or decrease of speech latencies within each task (i_4-i_1). Latency means of each task and of each interval as well as latency differences were defined as sVF features for prediction analysis.

Semantic distances were computed using two different approaches to ensure that the results of prediction analysis are not dependent on a specific semantic system. One of the semantic systems was a hierarchical structured lexical-semantic net of *GermaNet* [79] and *GermaNet-Pathfinder* [80]. Specifically, this lexical network is partitioned into various sets of semantic concepts (*synsets*) that are intertwined by semantic relations and create nodes. These synsets are related conceptually in different ways including, hypernymy, part-whole relations, entailment and causation, [79] leading to hierarchical-structured subcategories. *GermaNet-*

Pathfinder [80] provides different measurements [81] for the determination of how closely two nouns are related to each other. In this study, we selected a path-based measure which describes the relatedness between concepts. In detail, the path-based system takes the distance between two synset nodes and the longest possible shortest path between any two nodes in GermaNet into account.

$$sim(s_1, s_2) = \frac{MAXSHORTESTPATH - length(s_1, s_2)}{MAXSHORTESTPATH}$$

length(s_1,s_2) = shortest path between synset s1 and synset s2
MAXSHORTESTPATH = maximum of all shortest paths within GermaNet

Applying this formula, semantic relatedness is represented by values between 0 and 1. While closely related words lead to values approximating 1 (German Shepard x Labrador → sim=0.94), more distanced word pairs lead to smaller values (e.g. German Shepard x dolphin → sim=0.77).

The other semantic system that was used to determine semantic similarity between words was *DISCO* [82] applying a Word2Vec [83] approach. This system is based on co-occurrences in large text corpora. Specifically, this corpus contains 1.5 billion tokens including German Wikipedia entries, newspaper articles, parliamentary debates, movie subtitles and more. Each unique word is represented by a word vector and is part of the vector space. Within this vector space, word vectors are located based on shared common contexts building word embeddings. As in *GermaNet* [79], a high semantic similarity is represented by numbers approximating 1.

Each sVF task of the participants was analysed automatically using *GermaNet Pathfinder* [80] and *DISCO API* [82]. For our feature-set which was later used for the prediction analysis, two different types of semantic relations were extracted: (1) Sequential distance was computed across each consecutive word pair in order of the produced words. (2) Cumulative distance was computed over the entire task regardless of the order in which they appear within the task. As an output, the relatedness between each word-pair was extracted and the mean of all semantic relations within one task was calculated. In the case that *GermaNet* contained more than one

synset for one word, the synset with closest relatedness to the paired word was selected. Moreover, missing lexical entries in *GermaNet* or *DISCO* led to a deletion of the corresponding word pair. All semantic information, including means of sequential and cumulative distances of both systems (*GermaNet* and *DISCO*) were added as features to prediction analysis.

Altogether, 43 features were extracted from the sVF tasks containing information of sum of correct words, error types, speech latencies and semantic distances calculated with two different systems. A complete overview of VF feature scores is provided in the supplementary material (Table S2).

Machine learning analysis

In this study, we applied a machine approach using a cross-validation procedure. Here, just parts of the data are used to train the model while the other part is used to validate the model; i.e. EF scores were predicted in unseen participants which allows for generalization of results to a certain degree.

EF performance was predicted from sVF variables (*features*) applying supervised learning via random forests [84,85] (RF). The sVF features were used to predict each of the 68 EF scores (*targets*) in separate and independent analyses. Generally speaking, RF creates a "forest" of decision trees as weak learners by randomly sampling the features before learning each decision tree. The trees are used as an ensemble and the prediction of individual trees is averaged to get the final prediction [86]. In the present study 100 trees were used to compute prediction analysis. Previous work indicates that performance in the VF task is negatively related to age [87,88]. Moreover, sex was found to be associated with differential solving strategies in the VF task [89]. Likewise, a higher level of education was associated with better performance in VF tasks [87,90]. Therefore, data was transformed to z-scores and sex, age and education were regressed out from the sVF features within cross-validation. A 10-fold cross-validation procedure was performed for which the data set was randomly split into 10 sets, 9 of which were used for training while

287 the 10th set was held back and used to assess the prediction performance in previously unseen
288 data. Ten repetitions of the 10-fold cross-validation were performed and thus 100 prediction
289 models for each EF target were computed. Prediction performance was assessed by computing
290 the mean correlation (*Pearson*) between real and predicted values within cross-validation folds
291 and subsequently across all repetitions. EF targets which were predicted from sVF features at
292 a significance level of $p<0.01$ were considered *highly predictable* EF targets.
293 To compare the predictive power of the comprehensive and the classical feature set, the
294 prediction analysis was computed for classical sVF features, solely containing information from
295 sum scores of sVF tasks.
296 The sVF features which contributed most strongly to the prediction analyses of each *highly*
297 *predictable* EF target were identified. Feature importance was defined by the permutation of
298 out-of-bag predictor observations as implemented in *Matlab* [91]. The top five sVF features with
299 the highest feature performance were identified to further investigate the (non) linear
300 relationship of these sVF features with the respective EF performance. Here, rank correlations
301 (*Spearman*) of sVF features and EF test scores were calculated. Due to the high number of
302 extracted EF variables, only one highly significantly predicted EF test variable of each EF test
303 is presented to exemplarily demonstrate the complex relationship of sVF features and EF
304 performance. The selection of this representative EF variable was based on the test EF manuals
305 and previous literature describing specific main variables of each EF test.

Results

Prediction of EF variables from verbal fluency data

To investigate which EF targets were predictable from sVF features, we computed two independent prediction analyses. In the first analyses the full set of sVF features, including sum scores, errors, latencies and semantic relatedness was used (*Fig. 1*). The second analysis was performed with variables containing only information regarding the number of correctly produced items in each sVF task (*Fig. 2*). Both figures show the EF targets that were significantly predicted from sVF features at a significance level of $p<0.01$. Detailed results of all prediction analyses are given in the supplementary material (*Table S3)*.

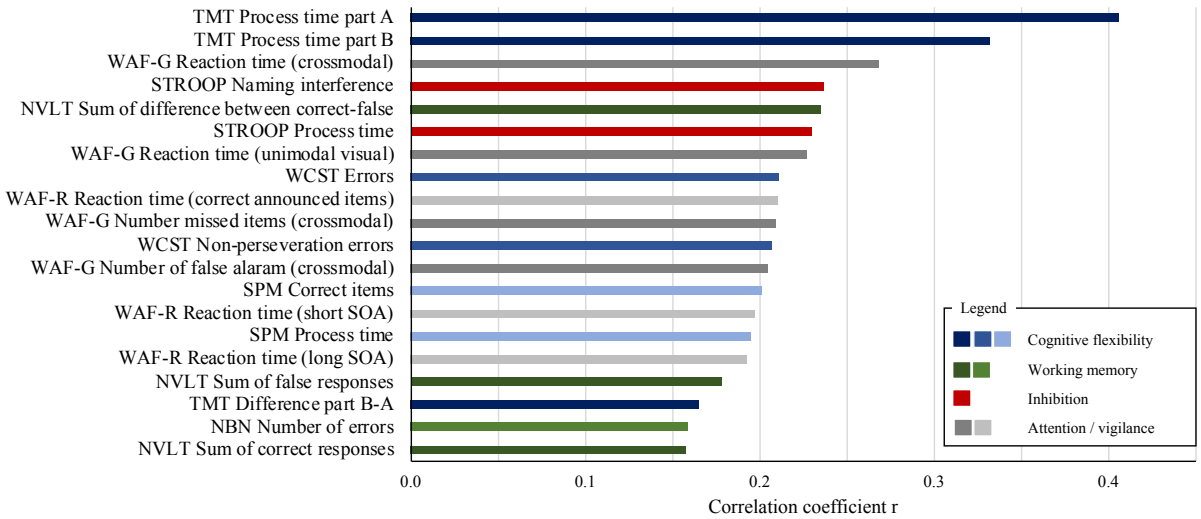

Figure 1: Correlation coefficients of true and predicted executive function variables computed with full feature set
Executive function variables were predicted based on 43 verbal fluency features. Results shown in this table illustrate executive function variables which could be predicted at p<0.01 from verbal fluency data; Colour groups indicate EF domains and colour gradients denote different EF tests within this EF domain;
NBN=N-back non-verbal; NVLT=Non-verbal Learning Test; SOA=Stimulus onset asynchrony; SPM=Raven´s Standard Progressive Matrices; STROOP=Stroop Test; TMT=Trail Making Test; WCST=Wisconsin Card Sorting Test; WAF-G=Divided attention; WAF-R=Spatial attention.

In sum, 20 EF targets, pertaining to 8 different EF tests and tapping into all subdomains of EFs, could be predicted significantly from the full feature set. With respect to cognitive flexibility, TMT, SPM and WCST were predicted from sVF data. The highest correlation between true and predicted values was identified in processing times of part A ($r=0.41$; $p=3.2e^{-10}$) and B ($r=0.33$; $p=2.6e^{-7}$) of TMT. While these results are primarily related to overall processing speed, an explicit relationship between sVF performance and cognitive flexibility was found in the

"*difference between part B-A*" of the TMT (r=0.17 *p*=0.007) as well as in the test results of SPM and WCST. Here, the number of correct items in the SPM (r=0.20; *p*=0.001) and different error types in the WCST revealed the complexity of cognitive requirements and planning ability in conducting the sVF task. With regards to tests assessing working memory capacity, two of three EF tests, namely NVLT (r=0.24; *p*=0.0002) and NBN (r=0.16; *p*=0.009) were predicted significantly. Beside EF targets referring to cognitive flexibility and working memory, the analysis also identified inhibition targets. Particularly, *naming interference* (r=0.24; *p*=0.0002) and processing time in STROOP (r=0.23; *p*=0.0003) were predicted.

Across all subdomains of EFs, variables displaying general processing speed and reaction times performance were detected. The role of attention and general processing speed is also represented in test results referring to divided and spatial attention. Here, seven targets of attention tests were predicted significantly. In general, tests from all EF subdomains were predicted significantly and no dominance of one specific subdomain was apparent.

The focus of this study was the predictive power of an advanced VF feature set. To compare the predictive power of the advanced features with the commonly used VF information, i.e. the sum of correctly produced words, an additional prediction analysis was computed using solely sum scores. Here, the sum scores of each sVF tasks as well as a total sum score across all three tests were included. In this analysis only six EF targets were predicted significantly (*Fig. 2*). Prediction performance was lower than in the analysis with full feature set and particularly targets of processing speed and reactions times were detected. In contrast to the first analysis, vigilance was predicted with missed items in CLOCK (r=0.16; *p*=0.009).

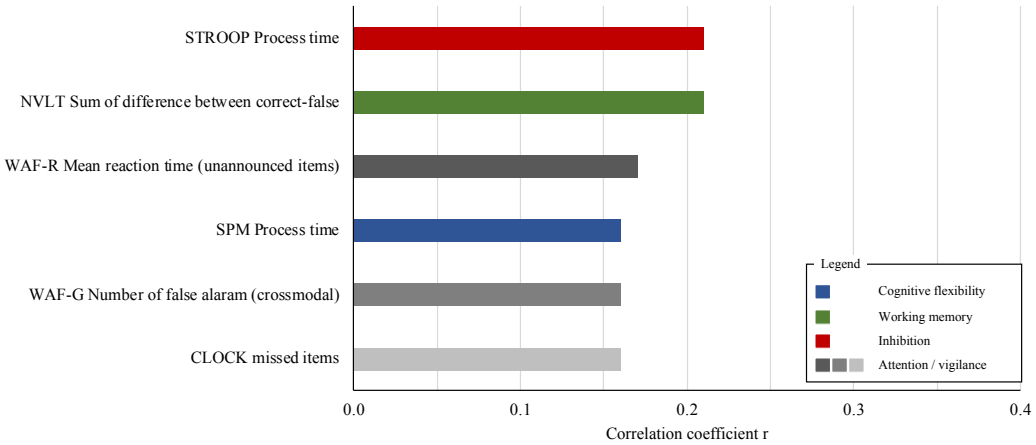

Figure 2: Correlation coefficients of true and predicted executive function variables computed with classical feature set
Executive function variables were predicted based on the sum scores of all 3 semantic verbal fluency tests as well as the total sum score across these 3 tests, which led to a total number of four verbal fluency features. Results shown in this table illustrate executive function variables which could be predicted with p<0.01 from verbal fluency data; Colour groups indicate EF domains and gradients denote different EF tests within this EF domain;
MACK=Mackworth Clock Test; NVLT=Non-verbal Learning Test; SPM=Raven's Standard Progressive Matrices; STROOP=Stroop Test; WAF-G=Divided attention; WAF-R=Spatial attention.

Impact of sVF features on prediction analysis

The impact of single sVF features on EF performance was quantified based on the feature importance scores of the prediction analysis. Due to the high number of significantly predicted EF targets, only one EF target for each of the significantly predicted EF tests is discussed in detail here. We focus on the main variables for the respective EF tests based on previous literature and the EF test manuals. For each of these, the five most important sVF features were extracted and correlations with the respective EF target were calculated (*Tab. 2-5*) to enable a comparison of present results with commonly used univariate analyses.

Across all EF domains, the most important sVF features for the prediction results included information about number of correctly produced words, error types, latencies and semantic distances. Out of these most predictive sVF features, some showed a significant correlation with the EF target ($p<0.05$), while others displayed a trend level significance ($p<0.1$) or no significant correlation at all. In the following, we assessed the top five sVF features that are related to the different EF subdomains of cognitive flexibility, working memory, inhibition as well as to attention. Due to the high number of EF scores that were predicted significantly from sVF features, one EF variable of each significantly predicted test is presented here. A complete

overview of the correlation matrix of all sVF features and significantly predicted EF scores is given in the supplementary material (*Table S4-S6*).

Table 2: Spearman correlations of five most important semantic verbal fluency (sVF) features with significantly predictable cognitive flexibility targets

	SPM – Correct items				TMT – Difference part B-A				WCST – Non-perseveration errors			
	Top 5 sVF features	r	p		Top 5 sVF features	r	p		Top 5 sVF features	r	p
1	Repetition error t_1	-0.05	0.44	1	Latencies 4th quarter t_1	-0.01	0.88	1	Repetition error t_3	-0.10	0.12
2	Latencies 2nd quarter t_1	-0.09	0.18	2	**Repetition error t_3**	**-0.12**	**0.08**	2	**Latencies 1st quarter t_1**	**0.17**	**0.01***
3	Category error t_3	-0.01	0.87	3	Latency difference t_1	-0.02	0.80	3	Category error t_3	0.01	0.85
4	**Correct words t_1**	**-0.12**	**0.08**	4	Category error t_3	-0.07	0.27	4	**Correct words t_2**	**-0.16**	**0.02***
5	**Cum. mean t_3**	**0.14**	**0.05***	5	Repetition error t_1	-0.01	0.80	5	**Total sum score t_1+t_2+t_3**	**-0.16**	**0.02***

1-5=Top five sVF features with regards to predictor performance based on feature importance; correlations with $p < 0.1$ are marked in bold; significant correlations ($p < 0.05$) are marked with *;
SPM=Raven´s Standard Progressive Matrices; TMT=Trail-Making Test; WCST=Wisconsin Card Sorting Test;
t_1=VF test (animals); t_2= VF test (jobs); t_3= Switching VF test (sports/fruits); Cum=cumulative

With regards to cognitive flexibility (*Tab.2*) 7/15 sVF features were related to errors participants produced within the sVF task. Repetition errors in simple and switching sVF tasks as well as category errors in the switching task were found to be important sVF features for predicting EF targets. Particularly, repetition and category errors were determined as highly relevant in predicting *TMT* performance. However, no significant (linear) correlation between errors and cognitive flexibility performance was found. In contrast, a linear relationship of sVF information and EF performance was shown for the number of correctly produced words. Here, significant correlations of correctly produced words and cognitive flexibility targets were primarily found in the *WCST*. Similar but not significant results were also found in the *SPM*. In all three significantly predicted EF tests (*SPM, TMT, WCST*) latencies within the sVF task$_1$ (animals) were identified as important sVF features but did not reveal correlations with EF targets except for latency patterns assessed in i_1. Here, longer speech breaks were shown to positively correlate with errors in WCST. With regards to semantic relatedness the cumulative mean within the sVF switching task (t_3), calculated with the hierarchical structured approach of *GermaNet*, was identified as a meaningful feature predicting *SPM* performance. Specifically, participants naming closely related words across both switching categories (sports and fruits) achieved better *SPM* targets.

Table 3: Spearman correlations of five most important semantic verbal fluency (sVF) features with significantly predictable working memory targets

	NBN – Errors				NVLT – Difference correct minus errors		
	Top 5 sVF features	r	p		Top 5 sVF features	r	p
1	Repetition error t_3	-0.01	0.91	1	**Latencies 4th quarter t_1**	**-0.12**	**0.08**
2	**Sequ. mean t_1**	**-0.19**	**0.00***	2	Category error t_3	-0.03	0.63
3	Category error t_3	-0.01	0.92	3	**Latency difference t_1**	**-0.12**	**0.08**
4	Repetition error t_1	0.11	0.11	4	Correct words t_1	0.07	0.28
5	**Cum. mean DIS t_3**	**-0.12**	**0.07**	5	Latency mean t_2	0.03	0.69

1-5=Top five VF features with regards to predictor performance based on feature importance; correlations with $p < 0.1$ are marked in bold; significant correlations ($p < 0.05$) are marked with *;
NBN=N-back non-verbal; NVLT=Non-verbal learning test
t_1=VF test (animals); t_2= VF test (jobs); t_3= Switching VF test (sports/fruits); Cum=cumulative; Sequ=sequential; DIS=semantic system *DISCO*

Within the EF domain of working memory, the *NBN* and *NVLT* were identified as highly predictable EF tests (*Tab.3*). Here, the sum of correctly produced words was selected as an important sVF feature less often than for cognitive flexibility tests and no significant correlation with EF target was found. Non-linear relationships of sVF features and working memory performance was additionally found for sVF features *errors* which were mainly important for predicting *NBN* performance. Among the five most important sVF features predicting NBN performance, the sequential as well as cumulative mean of the semantic relatedness were found to be highly relevant. Similar to results in cognitive flexibility tests (*Tab.1*), a smaller search space (r=-0.12 *p*=0.07) and closely related words (r=-0.19 *p*=0.005) led to better results in *NBN*. While semantic relatedness was particularly important for predicting errors in *NBN*, latencies were relevant for *NVLT* performance. Here, results indicated a relationship between smaller speech breaks in end of the sVF task and higher NVLT target (r=-0.12 *p*=0.08).

Table 4: Spearman correlations of five most important semantic verbal fluency (sVF) features with significantly predictable inhibition target

STROOP – Naming interference			
Top 5 sVF features		r	p
1	**Cum. mean t_2**	**-0.18**	**0.01***
2	Latency difference t_1	0.03	0.62
3	Latencies 4th quarter t_1	0.05	0.48
4	**Sequ. mean DIS cat_1 t_3**	**-0.15**	**0.03***
5	**Total sum score t_1+t_2+t_3**	**-0.22**	**0.00***

1-5=Top five VF features with regards to predictor performance based on feature importance; correlations with $p < 0.1$ are marked in bold; significant correlations ($p < 0.05$) are marked with *;
t_1=VF test (animals); t_2= VF test (jobs); t_3= Switching VF test (sports/fruits); Cum=cumulative; Sequ=sequential; DIS=semantic system *DISCO*

With respect to inhibition, naming interference in the *Stroop* test was predicted significantly. While error types were not selected as most important sVF features, the total sum score across all three sVF tests was determined as meaningful and revealed a significant correlation with *Stroop* performance (r=-0.22 p<0.001) (*Tab.4).* Important features for predicting *naming interference* performance were semantic relatedness and latencies. In particular, the searching space in t_2 represented by the cumulative mean was identified as highly important. These results indicate a better inhibition performance if participants searched for less distanced words (r=-0.18 p=0.01). Similar results were also found in sVF features of sequential relatedness. Searching for closely related words in the first category within the switching sVF task (cat_1 t_3) was related to better inhibitory performance (r=-0.15 p=0.03). Beside sematic relatedness and total sVF sum score, the analysis also points toward the relevance of latency patterns within the first sVF task (animals) for predicting inhibitory processes.

Table 5: Spearman correlations of five most important semantic verbal fluency (sVF) features with significantly predictable attention targets

WAF-G Reaction time crossmodal				WAF-R – Reaction time correctly announced			
Top 5 sVF features		r	p	Top 5 sVF features		r	p
1	Repetition error t_3	-0.09	0.17	1	Repetition error t_2	-0.01	0.93
2	Repetition error t_2	-0.03	0.70	2	Repetition error t_1	-0.09	0.17
3	**Latencies 1st quarter t_3**	**0.13**	**0.06**	3	Sequ. mean DIS t_1	0.04	0.61
4	Repetition error t_1	-0.05	0.45	4	Cum. mean DIS t_1	0.07	0.29
5	Cum. Mean t_1	-0.03	0.59	5	Repetition error t_3	0.01	0.88

1-5=Top five VF features with regards to predictor performance based on feature importance; correlations with $p < 0.1$ are marked in bold; significant correlations ($p < 0.05$) are marked with *;
WAF-G=divided attention test; WAF-R=spatial attention test;
t_1=VF test (animals); t_2= VF test (jobs); t_3= Switching VF test (sports/fruits); Cum=cumulative; Sequ=sequential; DIS=semantic system *DISCO*

448 Finally, we investigated sVF features in the prediction of attentional performance (*Tab.5*). Here,
449 the results demonstrate a predictive importance of repetition errors in simple as well as in
450 switching sVF tasks. The results revealed no significant correlation between number of errors
451 and attention performance. Latencies within the first quarter of the switching sVF task (t_3) were
452 selected as relevant for attention performance, indicating that a higher processing speed in the
453 beginning of the sVF task resulted in faster reaction times in the divided attention test. Similar
454 to previously reported results in other EF subdomains, semantic relatedness features in simple
455 sVF task (animals) were selected as meaningful variables for attention performance.
456 To sum up, across all subdomains of EFs, a variety of different types of sVF features, including
457 sum scores, error types, sematic relatedness and latencies showed high relevance for the
458 prediction of EF performance. Out of these, about one third showed significant or trend level
459 correlation with EF targets, while the remaining VF features that were identified as important
460 for prediction accuracy, did not show any linear relationship with the respective EF target.
461

Discussion

Main findings

This study aimed to investigate whether EF performance can be predicted from sVF tasks using Machine Learning methods. In a first step, we applied a RF approach to determine which EF tests could successfully be predicted from a wide range of VF information. Results of this machine learning analysis identified EF tests tapping into all subdomains of EFs. In total, 20 of 44 EF scores were predicted significantly when using the full set of sVF features which included errors, latencies and semantic distances.

Moreover, prediction results of the full sVF features set was compared to a classical feature set including only sum scores of sVF tasks, as commonly used in clinical settings. The comparison of these two approaches revealed a larger number of significantly predicted EF scores as well as higher prediction accuracy of the advanced feature set. Particularly for cognitive flexibility performance, the comprehensive feature set achieved a higher prediction accuracy as compared to the commonly used sum score evaluation. Thus, the present results clearly demonstrate the advantage of using more comprehensive sVF features over the sole use of sum scores, which to date still tend to be the most common measure used to asses sVF tasks. In a second step, we further investigated the concrete involvement of different types of sVF features to gain insights into the impact of specific VF aspects on EF performance. Results showed that all types of sVF features, i.e. sum scores, errors, latencies and semantic relatedness contributed to the prediction of EF. With regards to the different EF subdomains no dominance of specific VF types was detected. Moreover, the correlation analyses revealed that good sVF predictors do not necessarily correlate with the respective EF score.

The following section starts with a discussion of the influence of different sVF features on prediction results. Here, predictable EF tests within each subdomain are presented and the contributions of sVF features are interpreted. Additionally, the role of general processing speed

is addressed. Secondly, advantages of an elaborated VF feature set are delineated. In the end, limitations of this study are considered.

Sum scores

Summarizing scores of correctly produced items is the most commonly used way of evaluating VF tasks in the clinical and scientific context to date. The present study included separate sum scores for each sVF test as well as a total one across all sVF tasks. Results revealed the importance of sum score features for the prediction of cognitive flexibility, working memory and inhibition performance. In contrast, sum scores were not identified as important for predicting attention scores. Particularly sum scores resulting from t_1 (animals) as well as total sum scores revealed high feature importance. Furthermore, a positive linear relationship of relevant sVF sum scores and EF performance in the domains of inhibition and cognitive flexibility was found.

The findings from the present study can be directly linked to previous studies. In particular, Paula et al. [92] reported a positive correlation between cognitive flexibility performance, assessed with the TMT, and the sum of correct produced words in the switching task. With regards to working memory, another study found an association between the sum of correct produced items and working memory performance [43]. With respect to inhibition, our findings are also in line with multiple studies that demonstrated the positive linear relationship of inhibition performance and the total sum of words, assessed within the VF task, both in older [22] and young [93] adults.

Overall, based on previous literature and the results of the current study, sum scores were shown to contribute to the prediction results. In accordance with previous findings, this contribution appears to be based on a positive linear relationship of sum scores with EF performance.

Error types

When predicting EF test scores from sVF features, repetition and category errors were identified to mainly contribute to the prediction of cognitive flexibility, working memory and attention test result. Conversely, errors were not identified as important features for the prediction of inhibition scores. While both repetition and category errors were shown to be equally important for the prediction of cognitive flexibility and working memory, only repetition errors contributed to predicting attention performance. Importantly, in contrast to sum scores, most error features did not show a linear relationship with the respective EF test performance. The prediction results of TMT were the only ones to reveal a correlation trend, indicating that fewer repetition errors in sVF tasks are associated with better cognitive flexibility performance.

These findings partially contradict previous findings investigating the linear relationship between errors in the VF task and EF performance. Particularly, previous studies suggested that executive inhibitory dysfunction and reduced working memory performance lead to a higher number of perseveration errors in healthy participants [20,94]. Similar findings have also been reported in patients with brain damages [95] and schizophrenia [38]. In contrast, some studies did not find an increase in the number of perseveration errors in Parkinson´s patients compared to healthy controls [96].

Although in the present study repetition and category errors were shown to be important for the successful prediction of EF performance in all EF domains except for inhibitory processes, results revealed that a low number of produced errors does not necessarily result in better EF performance. Due to the importance of errors in prediction results and the non-linear relationship with EF performance, we assume that some participants adopt strategies where a higher number of errors is accepted in order to achieve a better score in the sVF task. Thus, successful EF performance does not necessarily go along with fewer errors.

Latencies

With regards to latency patterns our results revealed the importance of speech breaks for the prediction of all domains of EF as well as for attention scores. Latency patterns contributed differently to the prediction of different EF scores. Latency patterns during the first interval of the sVF task (i_1) were revealed as a meaningful feature for inhibitory processes, cognitive flexibility and attention performance. However, additional latency patterns, such as the mean of all latencies within each task and the progress of latencies (namely *latency differences*) also contributed to the prediction results. Interestingly, our results indicate an ambiguous relationship between latency patterns and EF test results. On the one hand, correlation analyses revealed some significant correlations between latency patterns and EF scores with, for example, longer speech breaks in i_1 were related to a higher amount of errors in the WCST assessing cognitive flexibility performance. On the other hand, most of the latency features did not show a linear relationship with EF performance.

To our knowledge, the relationship between speech breaks and EF performance in the context of VF has rarely been reported in previous literature, with existing studies tending to rather focus on unfilled pauses in free speech [56]. However, previous findings support our results with respect to the importance of speech breaks within the first interval of the VF task in that previous studies found a relationship between longer latencies in the beginning of the VF task and cognitive flexibility performance [39]. Moreover, other studies suggest that a decrease of speech latencies over the course of the VF task is related to the cluster patterns of the participants. While participants are assumed to produce clusters with high-frequent words in the beginning of the task, less frequent words are produced during the progression of the task leading to more switches and increased searching times [97].

In general, previous studies support the positive relationship between the duration of speech breaks and higher cognitive demands [56]. However, our results revealed mostly non-linear relationships between latencies and significantly predicted EF scores. This might suggest that

563 shorter speech breaks per se do not go along with better EF performance. Rather, we assume
564 that the heterogeneity of searching strategies, including processes such as clustering and
565 switching, lead to ambiguous latency patterns.

566
567 Semantic relatedness

568 Investigating the role of semantic relatedness between produced words within the sVF task, two
569 different semantic analysis systems were applied. On the one hand, a hierarchical approach was
570 used (*GermaNet*) [52]. On the other hand, an approach based on word embeddings was applied
571 (DISCO) [51]. The main goal of including both approaches was to assess as much diverse
572 semantic information as possible. Our results revealed that semantic relatedness measures from
573 both semantic systems contribute essentially to the prediction of all EF domains as well as to
574 attention performance. Although not all semantic features revealed a linear relationship with
575 EF performance, results indicate that searching for closely related words might be related to
576 stronger EF test results.

577 These findings are partially in line with previous studies which apply earlier approaches of
578 cluster and switching quantification to investigate the importance of switches in the sVF task
579 [92]. Authors have found a positive relationship between fewer switches and better cognitive
580 flexibility performance in healthy participants [92]. In contrast, other studies reported a decreased
581 number of switches in depressive patients with reduced cognitive flexibility [98]. Although the
582 present study did not differentiate between the two types of switches [42], the semantic systems
583 applied in this study [51,79] provided additional semantic distances which are similarly
584 interpretable. In detail, these semantic measurements also quantify semantic distances of
585 sequential and cumulative word pairs. Thus, a higher semantic mean in the present study can
586 be equated to a higher cluster size and less hard switches. However, the present study did not
587 aim to investigate such a fine-grained semantic approach as Troyer´s [41] approach but rather
588 strived to investigate the general importance of semantic distances within the sVF task.

589 In general, we assume that the production of semantically distanced words puts higher demands
590 on cognitive processes. However, for the sVF task, participants are asked to simply produce as
591 many words as possible, with no demands on the number of different subcategories these words
592 come from. Thus, producing closely related words and building high cluster sizes might
593 represent the most efficient strategy of successful EF performers.
594
595 Superiority of advanced sVF feature set
596 While the full feature set of sum scores, errors, latencies and semantic relatedness was applied
597 for the main analysis, we also predicted EF scores using sum score features only. Using the
598 sophisticated feature set, test variables from all EF domains as well as attention performance
599 and 8/14 EF tests were successfully predicted. While many of the predictable EF scores
600 contained general information of processing speed and reaction times, results also comprised
601 EF scores which are considered as characteristic variables for specific EF tests. For example,
602 TMT is represented by the *difference between part A-B* [99], Stroop by *naming interference* [72],
603 SPM by *correct items* [63] and WCST by *non-perseveration errors* [59], all of which were found to
604 be predictable EF scores.
605 In contrast, analysis with a classical sVF feature solely containing information of the sum
606 scores, predicted only 6/14 EF tests most of which were related to general processing speed
607 rather than to specific EF functions. Only one EF score of NVLT contained characteristic
608 information of working memory performance. EF scores representing cognitive flexibility and
609 inhibitory performance did not include information which are directly linked to EF performance
610 but rather related to general speed.
611 To our knowledge, so far, no other study has attempted combining different types of sVF
612 measurements to predict EF scores. However, previous research has demonstrated the
613 advantages of advanced approaches evaluating additional information over the sole use of the
614 total number of correctly produced words. For example, it was shown that the switching sVF

task, which was also used in the present study (t_3), contained more information of cognitive flexibility than simple VF tasks [92]. Our findings are also in line with another study investigating the digitalized evaluation of semantic relatedness with WordNet [50]. In particular, semantic relatedness was found to be highly associated with EFs and serve as an indicator for mild cognitive impairments which are difficult to detect with sum scores [100].

The comparison of prediction analysis with and without an extended set of sVF features mainly indicated that sum scores alone capture mostly working memory performance and attention scores. On the other hand, an advanced sVF feature set including sum scores, errors, latencies and sematic relatedness allows for the prediction of cognitive flexibility, working memory and inhibition performances as well as attention scores.

Investigating the relationship between the most important sVF features and EF performance in more detail, multiple non-linear relationships were detected. These findings highlight the advantages of machine learning approaches which are able to detect complex, non-linear relationships in addition to straightforward linear ones. Also, these approaches can take into account multivariate interactions between different VF features to reveal patterns which could not have been identified based on each single feature alone.

In general, our findings indicate that the use of a comprehensive set of VF features might have the potential to replace time-consuming and artificial EF tests. Due to the use of abstract symbols like numbers and letters, commonly used neuropsychological tests are criticized for their lacking ecological validity [11]. In contrast, producing words which are related to a specific category better represents daily needs and requirements of participants. Moreover, the lack of ecological validity might have influenced the correlations of the abstract EF test scores and the more natural sVF features. However, it remains open whether comprehensive sVF features may be even more helpful in clinical practice than commonly used EF test batteries.

Role of processing speed

In both analyses, variables which are not directly linked to EF performance but rather represent overall processing speed or reaction times, were predicted significantly. Similar findings were reported in our previous study predicting VF sum scores from EF tests variables [101]. The relationship of processing speed and sVF performance is also reported in other studies [88,102]. These authors suggest that processing time reflects general cognitive abilities such as intelligence to some extent [103] but may also be related to age [104] or personality traits such as extraversion [105]. Additionally, the presence of a time indication within some EF tests might facilitate processing speed similarly as in sVF tasks.

Limitations

Our results yielded insights into the involvement of EFs in the sVF task and highlighted the informative value of the sVF task to predict EF performance using a comprehensive feature set. Moreover, our results revealed complex and mostly non-linear relationships of VF features and EF performance. Hence, a detailed examination of individual differences in searching strategies might improve our understanding of which sVF patterns are related to higher EF performance in certain domains. As with all analyses of individual differences such research is dependent on large data sets comprising detailed information on EF and VF performance.

An additional consideration relates to the generalizability of our results. Ideally, our findings should be validated in a fully independent data set. To date, such a data set of sufficient size is not yet available. Hence, we applied a cross-validation approach within our sample. Here, the model was trained on some parts of the data while other parts of the data were held back. The model was then validated in the previously held back participants. This within-dataset validation represents the best alternative when a fully independently acquired dataset is not yet available.

Summary and outlook

Our study revealed insights into the advantages of an elaborated analysis of sVF tasks which successfully predicts EF performance. In comparison to the commonly used approach of evaluating sum scores of correctly produced words, we detected a lucid advantage of an extended feature analysis. In particular with regards to cognitive flexibility and inhibition our study demonstrated that an evaluation of sVF sum scores does not capture actual EF performance but rather assesses overall processing speed. Thus, we suggest the utilization of a comprehensive analysis of VF performance including features of error types, latencies and semantic distances. The present study applied primarily automated and digitalized methods ensuring a time-efficient and objective evaluation of VF performance. Further studies ought to develop a fully automated software tool integrating and further developing our feature set. Here, it would be highly interesting to also include features from the lexical VF task. A computerized toolbox allowing for an extensive assessment of VF could serve as a screening tool for EFs in a clinical diagnostic process as well as in a research context. Such a tool could include an audio system that records the speech of the patient and converts it into text. Subsequently, an automated software could be used to automatically determine a comprehensive set of VF features including sum scores, errors, latencies and semantic distances from the transcribed data. This can in turn result in a digitalized and quantified evaluation of the patient´s EFs compared to healthy controls based on VF performance, which can be then used by the clinician as part of the diagnostic process. Consequently, this toolbox could allow for higher ecological validity while also saving time in clinical routine.

However, we do not suggest that VF assessments will be able to fully substitute an initial extensive assessment of EFs with commonly used EF test batteries. We rather propose an extended and fully digitalized VF analysis as part of progress diagnostics in the form of a screening to assess EF performance in e.g. Parkinson´s disease or ADHD. Additionally, this screening-tool could be used in patients with predispositions of schizophrenia before

691 manifestation of clinical symptoms. Here, an advanced sVF analysis could provide insights into
692 subtle changes of EF performance. In the future, this work might contribute to an automated
693 digitalized speech analysis supporting clinicians in diagnostic processes.
694 Altogether, the present study demonstrated the predictive superiority of an extended VF feature
695 evaluation. Additionally, the results provided a first step towards an automated analysis of VF
696 serving as a predictor for EFs.

References

1. Friedman, N. P. & Miyake, A. Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex* **86**, 186–204 (2017).
2. Diamond, A. Executive functions. *Annu. Rev. Psychol.* **64**, 135–168 (2013).
3. Karr, J. E. *et al.* The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychol. Bull.* **144**, 1147–1185 (2018).
4. Kudlicka, A., Clare, L. & Hindle, J. V. Executive functions in Parkinson's disease: Systematic review and meta-analysis. *Mov. Disord.* **26**, 2305–2315 (2011).
5. Umarova, R. M. *et al.* Cognitive reserve impacts on disability and cognitive deficits in acute stroke. *J. Neurol.* **266**, 2495–2504 (2019).
6. Tavares, J. V. T. *et al.* Distinct profiles of neurocognitive function in unmedicated unipolar depression and bipolar II depression. *Biol. Psychiatry* **62**, 917–924 (2007).
7. Nigg, J. T., Blaskey, L. G., Huang-pollock, C. L. & Rappley, M. D. Neuropsychological Executive Functions and DSM-IV ADHD Subtypes. *J. Am. Acad. Child Adolesc. Psychiatry* **41**, 59–66 (2002).
8. Stuss, D. T. & Alexander, M. P. Executive functions and the frontal lobes: a conceptual view. *Psychol. Res.* **63**, 289–298 (2000).
9. Fine, E. M. & Delis, D. C. Delis–Kaplan Executive Functioning System. in *Encyclopedia of Clinical Neuropsychology* (eds. Kreutzer, J. S., DeLuca, J. & Caplan, B.) 796–801 (Springer New York, 2011). doi:10.1007/978-0-387-79948-3_1539.
10. *Wiener Testsystem*. (SCHUHFRIED GmbH, 2016).
11. Chan, R., Shum, D., Toulopoulou, T. & Chen, E. Assessment of executive functions: Review of instruments and identification of critical issues. *Arch. Clin. Neuropsychol.* **23**, 201–216 (2008).
12. Baddeley, A. Working memory and language: an overview. *J. Commun. Disord.* **36**, 189–208 (2003).
13. Levelt, W. J. Accessing words in speech production: stages, processes and representations. *Cognition* **42**, 1–22 (1992).
14. Ho, A. K. *et al.* Verbal fluency in Huntington's disease: a longitudinal analysis of phonemic and semantic clustering and switching. *Neuropsychologia* **40**, 1277–1284 (2002).
15. Canning, S. J. D., Leach, L., Stuss, D., Ngo, L. & Black, S. E. Diagnostic utility of abbreviated fluency measures in Alzheimer disease and vascular dementia. *Neurology* **62**, 556–562 (2004).
16. Pakhomov, S. V. S. & Hemmy, L. S. A computational linguistic measure of clustering behavior on semantic verbal fluency task predicts risk of future dementia in the Nun Study. *Cortex* **55**, 97–106 (2014).
17. Weiss, E. M. *et al.* Sex differences in clustering and switching in verbal fluency tasks. *J. Int. Neuropsychol. Soc.* **12**, (2006).
18. Nikolai, T. *et al.* Semantic verbal fluency impairment is detectable in patients with subjective cognitive decline. *Appl. Neuropsychol. Adult* **25**, 448–457 (2018).
19. Gonçalves, H. A. *et al.* Clustering and switching in unconstrained, phonemic and semantic verbal fluency: the role of age and school type. *J. Cogn. Psychol.* **29**, 670–690 (2017).
20. Azuma, T. Working Memory and Perseveration in Verbal Fluency. *Neuropsychology* **18**, 69–77 (2004).
21. Rosen, V. M. & Engle, R. W. The role of working memory capacity in retrieval. *J. Exp. Psychol. Gen.* **126**, 211–227 (1997).
22. Fisk, J. E. & Sharp, C. A. Age-Related Impairment in Executive Functioning: Updating, Inhibition, Shifting, and Access. *J. Clin. Exp. Neuropsychol.* **26**, 874–890 (2004).
23. Shao, Z., Janse, E., Visser, K. & Meyer, A. S. What do verbal fluency tasks measure?

Predictors of verbal fluency performance in older adults. *Front. Psychol.* **5**, (2014).
24. Whiteside, D. M. *et al.* Verbal Fluency: Language or Executive Function Measure? *Appl. Neuropsychol. Adult* **23**, 29–34 (2016).
25. Robinson, G., Shallice, T., Bozzali, M. & Cipolotti, L. The differing roles of the frontal cortex in fluency tests. *Brain* **135**, 2202–2214 (2012).
26. Hidalgo-Lopez, E. & Pletzer, B. Interactive Effects of Dopamine Baseline Levels and Cycle Phase on Executive Functions: The Role of Progesterone. *Front. Neurosci.* **11**, (2017).
27. Hedden, T. & Yoon, C. Individual differences in executive processing predict susceptibility to interference in verbal working memory. *Neuropsychology* **20**, 511–528 (2006).
28. Fournier-Vicente, S., Larigauderie, P. & Gaonac'h, D. More dissociations and interactions within central executive functioning: A comprehensive latent-variable analysis. *Acta Psychol. (Amst.)* **129**, 32–48 (2008).
29. Benjamin, M. J., Cifelli, A., Garrard, P., Caine, D. & Jones, F. W. The role of working memory and verbal fluency in autobiographical memory in early Alzheimer's disease and matched controls. *Neuropsychologia* **78**, 115–121 (2015).
30. Patra, A., Bose, A. & Marinis, T. Performance difference in verbal fluency in bilingual and monolingual speakers. *Biling. Lang. Cogn.* 1–15 (2019) doi:10.1017/S1366728918001098.
31. Clark, D. G. *et al.* Novel verbal fluency scores and structural brain imaging for prediction of cognitive outcome in mild cognitive impairment. *Alzheimers Dement. Diagn. Assess. Dis. Monit.* **2**, 113–122 (2016).
32. Zhu, Z., Novikova, J. & Rudzicz, F. Detecting cognitive impairments by agreeing on interpretations of linguistic features. *ArXiv180806570 Cs* (2019).
33. Cummins, N., Sethu, V., Epps, J. & Krajewski, J. Relevance vector machine for depression prediction. in *INTERSPEECH* (2015).
34. Bedi, G. *et al.* Automated analysis of free speech predicts psychosis onset in high-risk youths. *Npj Schizophr.* **1**, 1–7 (2015).
35. van den Berg, E., Jiskoot, L. C., Grosveld, M. J. H., van Swieten, J. C. & Papma, J. M. Qualitative Assessment of Verbal Fluency Performance in Frontotemporal Dementia. *Dement. Geriatr. Cogn. Disord.* **44**, 35–44 (2017).
36. Ruff, R. M., Light, R., Parker, S. B. & Levin, H. S. Benton Controlled Oral Word Association Test: reliability and updated norms. *Arch. Clin. Neuropsychol. Off. J. Natl. Acad. Neuropsychol.* **11**, 329–338 (1996).
37. Aschenbrenner, S., Tucha, O. & Lange, K. W. *Regensburger Wortflüssigkeits-Test: RWT*. (Hogrefe, Verlag für Psychologie, Göttingen).
38. Galaverna, F., Bueno, A. M., Morra, C. A., Roca, M. & Torralva, T. Analysis of errors in verbal fluency tasks in patients with chronic schizophrenia. *Eur. J. Psychiatry* **30**, 305–320 (2016).
39. Wolters, M. K., Kim, N., Kim, J.-H., MacPherson, S. E. & Park, J. C. Prosodic and Linguistic Analysis of Semantic Fluency Data: A Window into Speech Production and Cognition. in 2085–2089 (2016). doi:10.21437/Interspeech.2016-420.
40. Pakhomov, S. V. S., Eberly, L. & Knopman, D. Characterizing cognitive performance in a large longitudinal study of aging with computerized semantic indices of verbal fluency. *Neuropsychologia* **89**, 42–56 (2016).
41. 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).
42. Abwender, D. A., Swan, J., Bowerman, J. & Connolly, S. Qualitative Analysis of Verbal Fluency Output: Review and Comparison of Several Scoring Methods. *Assessment* (2001) doi:10.1177/107319110100800308.

43. Unsworth, N., Spillers, G. J. & Brewer, G. A. Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance. *Q. J. Exp. Psychol.* **64**, 447–466 (2011).
44. 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).
45. Price, S. E. *et al.* Semantic verbal fluency strategies in amnestic mild cognitive impairment. *Neuropsychology* **26**, 490–497 (2012).
46. Rich, J. B., Troyer, A. K., Bylsma, F. W. & Brandt, J. Longitudinal analysis of phonemic clustering and switching during word-list generation in Huntington's disease. *Neuropsychology* **13**, 525–531 (1999).
47. 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).
48. *Handbook of latent semantic analysis*. xii, 532 (Lawrence Erlbaum Associates Publishers, 2007).
49. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. Distributed Representations of Words and Phrases and their Compositionality. *ArXiv13104546 Cs Stat* (2013).
50. *WordNet: An Electronic Lexical Database*. (The MIT Press, 1998). doi:10.7551/mitpress/7287.001.0001.
51. Kolb, P. DISCO: A Multilingual Database of Distributionally Similar Words. 8.
52. Henrich, V. & Hinrichs, E. Determining Immediate Constituents of Compounds in GermaNet. in *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011* 420–426 (Association for Computational Linguistics, 2011).
53. Jurafsky, D. & Martin, J. H. *Speech and Language Processing (2nd Edition)*. (Prentice-Hall, Inc., 2009).
54. Pauselli, L. *et al.* Computational linguistic analysis applied to a semantic fluency task to measure derailment and tangentiality in schizophrenia. *Psychiatry Res.* **263**, 74–79 (2018).
55. Clark, H. H. Managing problems in speaking. *Speech Commun.* **15**, 243–250 (1994).
56. Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F. & Brennan, S. E. Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Lang. Speech* **44**, 123–147 (2001).
57. *Wiener Testsystem*. (SCHUHFRIED GmbH, 2016).
58. Stoet, G. PsyToolkit: A software package for programming psychological experiments using Linux. *Behav. Res. Methods* **42**, 1096–1104 (2010).
59. Bowden, S. C. *et al.* The Reliability and Internal Validity of the Wisconsin Card Sorting Test. *Neuropsychol. Rehabil.* **8**, 243–254 (1998).
60. Cohen, A.-L., Bayer, U. C., Jaudas, A. & Gollwitzer, P. M. Self-regulatory strategy and executive control: implementation intentions modulate task switching and Simon task performance. *Psychol. Res.* **72**, 12 (2006).
61. Mackworth, N. H. The breakdown of vigilance during prolonged visual search. *Q. J. Exp. Psychol.* **1**, 6–21 (1948).
62. Reitan, R. M. Validity of the Trail Making Test as an Indicator of Organic Brain Damage. *Percept. Mot. Skills* **8**, 271–276 (1958).
63. Raven, J. C., Raven, J. & Court, J. H. *SPM Manual (Deutsche Bearbeitung und Normierung von St. Bulheller und H. Häcker)*. (Swets & Zeitlinger B.V.).
64. Grant, D. A. & Berg, E. A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *J. Exp. Psychol.* **38**, 404–411 (1948).
65. Kaller, C. P., Unterrainer, J. M. & Stahl, C. Assessing planning ability with the Tower

of London task: psychometric properties of a structurally balanced problem set. *Psychol. Assess.* **24**, 46–53 (2012).
66. Meiran, N. Reconfiguration of processing mode to task performance. *J. Exp. Psychol. Learn. Mem. Cogn.* **22**, 1423–1442 (1996).
67. Schellig, D., Schuri, U. & Arendasy, M. *NBN- NBACK-nonverbal*. (SCHUHFRIED GmbH, 2009).
68. Sturm, W. & Willmes, K. *NVLT Non-Verbal Learning Test*. (SCHUHFRIED GmbH, 2016).
69. Schellig, D. & Hättig, H. A. Die Bestimmung der visuellen Merkspanne mit dem Block-Board. *Z. Für Neuropsychol.* **4**, 104–112 (1993).
70. Kaiser, S., Aschenbrenner, S., Pfüller, U., Roesch-Ely, D. & Weisbrod, M. *Response inhibition*. (SCHUHFRIED GmbH, 2016).
71. Simon, J. R. & Wolf, J. D. Choice reaction time as a function of angular stimulus-response correspondence and age. in (1963). doi:10.1080/00140136308930679.
72. Schuhfried, G. *Interferenz nach Stroop*. (SCHUHFRIED GmbH, 2016).
73. Sturm, W. *Wahrnehmungs- und Aufmerksamkeitsfunktionen: Geteilte Aufmerksamkeiten*. (SCHUHFRIED GmbH, 2016).
74. Aschenbrenner, Steffen, Tucha, Oliver & Lange, Klaus W. *Regensburger Wortflüssigkeits-Test: RWT*. (Hogrefe, Verlag für Psychologie, Göttingen).
75. Benton, A. L., Hamsher, S. K. & Sivan, A. B. *Multilingual aplasia examination*. (1983).
76. Ruff, R. M., Light, R., Parker, S. B. & Levin, H. S. Benton Controlled Oral Word Association Test: reliability and updated norms. *Arch. Clin. Neuropsychol. Off. J. Natl. Acad. Neuropsychol.* **11**, 329–338 (1996).
77. *Presentation software*. (Neurobehavioral Systems, 2018).
78. Boersma, P. Praat, a system for doing phonetics by computer. *Glot Int.* **5**, (2002).
79. Henrich, V. & Hinrichs, E. Determining Immediate Constituents of Compounds in GermaNet. in *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011* 420–426 (Association for Computational Linguistics, 2011).
80. Cramer, I. M. & Finthammer, M. Tools for exploring GermaNet in the context of cl-teaching. in *KONVENS* (2008). doi:10.1515/9783110211818.3.195.
81. Barsukova, A. *et al. Tutorial: Semantic Relatedness API for GermaNet*. (University of Tübingen; Department of General and Computational Linguistics, 2018).
82. Kolb, P. Experiments on the difference between semantic similarity and relatedness. 8.
83. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. Distributed Representations of Words and Phrases and their Compositionality. *ArXiv13104546 Cs Stat* (2013).
84. Tin Kam Ho. Random decision forests. in *Proceedings of 3rd International Conference on Document Analysis and Recognition* vol. 1 278–282 vol.1 (1995).
85. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
86. Denil, M., Matheson, D. & De Freitas, N. Narrowing the Gap: Random Forests in Theory and in Practice. in *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32* I-665-I–673 (JMLR.org, 2014).
87. Tombaugh, T. N., Kozak, J. & Rees, L. Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Arch. Clin. Neuropsychol. Off. J. Natl. Acad. Neuropsychol.* **14**, 167–177 (1999).
88. Elgamal, S. A., Roy, E. A. & Sharratt, M. T. Age and Verbal Fluency: The Mediating Effect of Speed of Processing. *Can. Geriatr. J. CGJ* **14**, 66–72 (2011).
89. Lanting, S., Haugrud, N. & Crossley, M. The effect of age and sex on clustering and switching during speeded verbal fluency tasks. *J. Int. Neuropsychol. Soc. JINS* **15**, 196–204 (2009).

90. Obeso, I., Casabona, E., Bringas, M. L., Alvarez, L. & Jahanshahi, M. Semantic and phonemic verbal fluency in Parkinson's disease: Influence of clinical and demographic variables. *Behav. Neurol.* **25**, 111–118 (2012).
91. Mathworks. Predictor importance estimates by permutation of out-of-bag predictor observations for random forest of regression trees - MATLAB. https://www.mathworks.com/help/stats/regressionbaggedensemble.oobpermutedpredictorimportance.html.
92. Paula, J. J. de, Paiva, G. C. de C. & Costa, D. de S. Use of a modified version of the switching verbal fluency test for the assessment of cognitive flexibility. *Dement. Neuropsychol.* **9**, 258–264 (2015).
93. Ardila, A., Galeano, L. M. & Rosselli, M. Toward a Model of Neuropsychological Activity. *Neuropsychol. Rev.* **8**, 171–190 (1998).
94. Carr, M., Saint-Onge, K., Blanchette-Carrière, C., Paquette, T. & Nielsen, T. Elevated perseveration errors on a verbal fluency task in frequent nightmare recallers: a replication. *J. Sleep Res.* **27**, e12644 (2018).
95. Fischer-Baum, S., Miozzo, M., Laiacona, M. & Capitani, E. Perseveration during verbal fluency in traumatic brain injury reflects impairments in working memory. *Neuropsychology* **30**, 791–799 (2016).
96. Suhr, J. A. & Jones, R. D. Letter and semantic fluency in Alzheimer's, Huntington's, and Parkinson's dementias. *Arch. Clin. Neuropsychol. Off. J. Natl. Acad. Neuropsychol.* **13**, 447–454 (1998).
97. Raboutet, C. *et al.* Performance on a semantic verbal fluency task across time: Dissociation between clustering, switching, and categorical exploitation processes. *J. Clin. Exp. Neuropsychol.* **32**, 268–280 (2010).
98. Fossati, P., Bastard Guillaume, L., Ergis, A.-M. & Allilaire, J.-F. Qualitative analysis of verbal fluency in depression. *Psychiatry Res.* **117**, 17–24 (2003).
99. Kortte, K. B., Horner, M. D. & Windham, W. K. The trail making test, part B: cognitive flexibility or ability to maintain set? *Appl. Neuropsychol.* **9**, 106–109 (2002).
100. Pakhomov, S. V. S., Hemmy, L. S. & Lim, K. O. Automated semantic indices related to cognitive function and rate of cognitive decline. *Neuropsychologia* **50**, 2165–2175 (2012).
101. Amunts, J., Camilleri, J. A., Eickhoff, S. B., Heim, S. & Weis, S. Executive functions predict verbal fluency scores in healthy participants. *Sci. Rep.* **10**, 11141 (2020).
102. Brébion, G. *et al.* Verbal fluency in male and female schizophrenia patients: Different patterns of association with processing speed, working memory span, and clinical symptoms. *Neuropsychology* **32**, 65–76 (2018).
103. Kail, R. & Salthouse, T. A. Processing speed as a mental capacity. *Acta Psychol. (Amst.)* **86**, 199–225 (1994).
104. Sliwinski, M. & Buschke, H. Cross-sectional and longitudinal relationships among age, cognition, and processing speed. *Psychol. Aging* **14**, 18–33 (1999).
105. Pearman, A. Basic cognition in adulthood: Combined effects of sex and personality. *Personal. Individ. Differ.* **47**, 357–362 (2009).

Acknowledgements

This research was supported by The Deutsche Forschungsgemeinschaft (DFG, EI 816/11-1), the National Institute of Mental Health (R01-MH074457), the Helmholtz Portfolio Theme "Supercomputing and Modeling for the Human Brain" and the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 720270 (HBP SGA1) 785907 (HBP SGA2).

We are very grateful to Lisa Mochalski, Natalie Schlothauer and Hannah Hensen for their help with testing participants. We also thank Tobias Kadelka for his help with the automated digitalization of the VF evaluation.

Author Contributions

The study was designed by all authors of the manuscript. In particular, the executive functions test selection was mainly supported by J.A.C. while S.H expertise mainly contributed to the discussion of speech-related topics. Data collection, data preprocessing and generation of VF features was done by J.A. Data analysis was mainly driven by S.W, K.R.P., J.A.C., J.A. and S.B.E.; Here, K.R.P. knowledge of different machine learning approaches and their implementation in *Matlab* was essential. Moreover, S.W. and S.B.E. contributed their knowledge of advanced data science. G.P. contributed contents relating to the psychiatric context as well as the influences of processing speed and intelligence on cognitive performance in psychiatric patients. The manuscript was mainly written by J.A., supervised throughout the writing process by S.W. and J.A.C. who provided improvements for structuring the manuscript as well as improving the wording. Moreover, S.B.E., K.R.P., S.H. and G.P. provided detailed feedback on the overall manuscript.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary Table S1: Test variables of executive function tests

Test	Abbreviation	Variable	Mean	SD
		COGNITIVE FLEXIBILITY		
Trail Making test	TMT	Process time partA	17.11	3.17
		Process time partB	24.70	7.51
		Diference part B-A [seconds]	7.59	6.30
		Quotient B/A	1.45	0.36
		Errors part A	0.06	0.28
		Errors part B	0.65	1.00
Raven´s Standard Progressive Matrices	SPM	Correct items	27.82	3.31
		process time	629.56	157.94
Wisconsin Card Sorting Test	WCST	Number of errors	12.68	6.17
		Number of perseveration errors	7.96	3.49
		Number of errors (non perseveration)	4.72	3.29
		Timeouts	0.52	1.05
Tower of London	TOL	Planning ability	7.52	2.19
		Number of correct responses	10.52	1.61
		Changed his/her mind, self-correction	1.82	2.08
		Choice of wrong pole	0.50	1.16
		Choice of blocked pole	1.03	2.11
		Choice of impossible position	0.79	2.34
Cued Task Switching	SWITCH	Number of errors	2.53	3.34
		Switch costs (reaction time switch tasks - reaction time in non-switch tasks) [seconds]	0.05	0.08
		Timeouts	0.28	0.78
		Errors of items which are incongruent	2.06	2.41
		WORKING MEMORY		
N-back non verbal	NBN	Correct items	8.36	2.94
		Number of commission errors	5.64	2.94
		Number of errors	8.86	7.89
		Mean reaction tine of correct items [seconds]	0.78	0.17
		Mean reaction time of errors [seconds]	0.87	0.24
Non-verbal learning test	NVLT	Sum of correct responses	32.09	5.03
		Sum of false responses	12.57	7.93
		Sum of difference between correct minus false responses	19.52	7.85
		Process time	119.16	33.97
Corsi Block Tapping Test	CORSI	Block span	5.66	1.09
		Correct items	10.59	2.81
		False items	4.59	1.31
		Missed items	0.04	0.19
		Sequency errors	2.59	1.27
		INHIBITION		
Stop Signal Task	INHIB	Reaction time [seconds]	0.51	0.09
		Mean stop signal delay [seconds]	0.30	0.06
		Stop signal reaction time [seconds]	0.21	0.07
		Number of commission errors	14.37	6.20
		Number of ommission errors	1.29	1.83
Simon Task	SIMON	Reaction time difference (reaction time incongruent - reaction time congruent items) [seconds]	0.03	0.04
		Number of errors in compatible items	2.85	11.17
		Number of errors in incompatible items	5.15	12.46
Stroop Test	STROOP	Reading interference [seconds]	0.14	0.08
		Naming interference [seconds]	0.13	0.08
		Interference-difference [seconds]	-0.05	0.12
		Number of false reactions (reading-baseline)	2.14	2.52
		Number of false reactions (naming-baseline)	2.67	2.47
		Number of false reactions (reading-interference)	3.36	6.60
		Number of false reactions (naming-interference)	3.52	3.19
		Process time	404.96	61.21
		ATTENTION / VIGILANCE		
Divided attention test	WAF-G	Number of missed items (unimodall visual)	2.08	3.10
		Number of false alarm (unimodal visual)	3.30	3.49
		Mean reaction time (unimodal visual) [ms]	445.70	89.16
		Number of missed items (crossmodal visual/auditive)	2.76	2.95
		Number of false alarm (crossmodal visual/auditive)	3.10	4.90
		Mean reaction time (crossmodal) [ms]	457.16	110.11
Spatial attention test	WAF-R	Mean reaction time (unannounced items) [ms]	379.84	47.69
		Number of missed items (correct announced items)	0.95	1.55
		Mean reaction time (correct announced items) [ms]	306.51	47.97
		Number of missed items (wrong announced items)	0.21	0.55
		Mean reaction time (wrong announced items) [ms]	339.46	52.46
		Mean reaction time (short SOA) [ms]	347.63	44.23
		Mean reaction time (long SOA) [ms]	341.17	44.24
		Number of errors	3.58	3.34
Mackworth Clock Test	MACK	Number of missed jumps	7.88	4.93
		Number of false alarms	3.15	3.78

Supplementary Table S2: Overview of verbal fluency task results

Variable	Description	Mean	SD
Correct words t1+t2+t3	Sum of all correct produced words in t1+t2+t3	84.83	15.54
	Verbal fluency task 1 (animals)		
Correct words t1	Sum of correct produced words in task 1	36.73	8.24
Repetition error t1	Repetition errors in task 1	0.60	0.89
Latency mean t1	Mean of speech breaks in task 1	2.88	0.86
Latencies 1st quarter t1	Mean of speech breaks in seconds 0-30 (i1)	1.41	0.48
Latencies 2nd quarter t1	Mean of speech breaks in seconds 31-60 (i2)	3.20	1.94
Latencies 3rd quarter t1	Mean of speech breaks in seconds 61-90 (i3)	4.23	2.71
Latencies 4th quarter t1	Mean of speech breaks in seconds 91-120 (i4)	6.48	6.51
Latency difference t1	Progress of speech breaks (i4-i1) in task 1	5.07	6.50
Sequential mean t1	Semantic mean of all sequential word pairs in task 1; computed with GermaNet (hierachical)	0.83	0.02
Cumulative mean t1	Semantic mean of all possible word pairs (cumulative) in task 1; computed with GermaNet (hierarchical)	0.80	0.04
Sequential mean t1 DIS	Semantic mean of all sequential word pairs in task 1; computed with DISCO (Word2Vec)	0.45	0.05
Cumulative mean t1 DIS	Cumulative mean of all sequential word pairs in task 1; computed with DISCO (Word2Vec)	0.36	0.03
	Verbal fluency task 2 (jobs)		
Correct words t2	Sum of correct produced words in task 2	26.08	6.65
Repetition error t2	Repetition errors in task 2	0.39	0.75
Latency mean t2	Mean of speech breaks in task 2	3.91	1.30
Latencies 1st quarter t2	Mean of speech breaks in seconds 0-30 (i1)	2.17	0.80
Latencies 2nd quarter t2	Mean of speech breaks in seconds 31-60 (i2)	4.06	1.95
Latencies 3rd quarter t2	Mean of speech breaks in seconds 61-90 (i3)	6.32	5.16
Latencies 4th quarter t2	Mean of speech breaks in seconds 91-120 (i4)	6.85	6.75
Latency difference t2	Progress of speech breaks (i4-i1) in task 2	4.68	6.69
Sequential mean t2	Semantic mean of all sequential word pairs in task 2; computed with GermaNet (hierarchical)	0.85	0.02
Cumulative mean t2	Semantic mean of all possible word pairs (cumulative) in task 2; computed with GermaNet (hierarchical)	0.82	0.02
Sequential mean t2 DIS	Semantic mean of all sequential word pairs in task 2; computed with DISCO (Word2Vec)	0.39	0.08
Cumulative mean t2 DIS	Semantic mean of all possible word pairs (cumulative) in task 2; computed with DISCO (Word2Vec)	0.27	0.03
	Verbal fluency task 3 (sports / fruits)		
Correct words t3	Sum of correct produced words in task 3	22.01	4.38
Repetition error t3	Repetition errors in task 3	0.20	0.48
Category error t3	Category errors in task 3	0.10	0.39
Latency mean t3	Mean of speech breaks in task 3	4.64	1.27
Latencies 1st quarter t3	Mean of speech breaks in seconds 0-30 (i1)	2.29	0.66
Latencies 2nd quarter t3	Mean of speech breaks in seconds 31-60 (i2)	5.16	2.15
Latencies 3rd quarter t3	Mean of speech breaks in seconds 61-90 (i3)	7.15	4.15
Latencies 4th quarter t3	Mean of speech breaks in seconds 91-120 (i4)	9.33	8.64
Latency difference t3	Progress of speech breaks (i4-i1) in task 3	7.04	8.59
Sequential mean t3	Semantic mean of all sequential word pairs in task 3; computed with GermaNet (hierarchical)	0.60	0.02
Cumulative mean t3	Semantic mean of all possible word pairs (cumulative) in task 3; computed with GermaNet (hierarchical)	0.74	0.02
Sequential mean t3 DIS	Semantic mean of all sequential word pairs in task 3; computed with DISCO (Word2Vec)	0.08	0.08
Cumulative mean t3 DIS	Semantic mean of all possible word pairs (cumulative) in task 3; computed with DISCO (Word2Vec)	0.23	0.02
Sequential mean cat1 t3	Semantic mean of all sequential word pairs in category 1 (sports) of switching task; computed with GermaNet (hierarchical)	0.89	0.03
Sequential mean cat2 t3	Semantic mean of all sequential word pairs in category 2 (fruits) of switching task; computed with GermaNet (hierarchical)	0.90	0.02
Sequential mean cat1 t3 DIS	Semantic mean of all sequential word pairs in category 1 (sports) of switching task; computed with DISCO (Word2Vec)	0.40	0.07
Sequential mean cat2 t3 DIS	Semantic mean of all sequential word pairs in category 2 (fruits) of switching task; computed with DISCO (Word2Vec)	0.52	0.05
Switch coefficient	relationship of simple and switching tasks - switching coefficient sum3/((sum1+sum2)/2))	-9.40	5.78
Correct words t1+t2+t3	Sum of all correct produced words in t1+t2+t3	84.83	15.54

Supplementary Table S3: Prediction results of executive function targets

			Full feature set		Reduced feature set	
Test	**Abbreviation**	**Variable**	**r**	**p**	**r**	**p**
	COGNITIVE FLEXIBILITY					
Trail Making test	TMT	Process time partA	0.41	**0.000**	0.11	0.050
		Process time partB	0.33	**0.000**	0.14	0.018
		Diference part B-A [seconds]	0.17	**0.007**	0.09	0.083
		Quotient B/A	0.11	0.055	0.07	0.146
		Errors part A	0.00	0.487	0.05	0.209
		Errors part B	0.07	0.136	0.05	0.222
Raven´s Standard Progressive Matrices	SPM	Correct items	0.20	**0.001**	0.04	0.295
		process time	0.19	**0.002**	0.16	**0.007**
Wisconsin Card Sorting Test	WCST	Number of errors	0.21	**0.001**	0.06	0.198
		Number of perseveration errors	0.11	0.045	0.02	0.406
		Number of errors (non perseveration)	0.21	**0.001**	0.06	0.189
		Timeouts	0.03	0.318	0.08	0.105
Tower of London	TOL	Planning ability	0.13	0.029	0.02	0.357
		Number of correct responses	-0.03	0.664	0.04	0.252
		Changed his/her mind, self-correction	0.06	0.171	-0.08	0.894
		Choice of wrong pole	0.07	0.137	-0.06	0.800
		Choice of blocked pole	0.00	0.526	0.01	0.432
		Choice of impossible position	0.02	0.401	0.01	0.428
Cued Task Switching	SWITCH	Number of errors	0.11	0.055	-0.06	0.826
		Switch costs (reaction time switch tasks - reaction time in non-switch tasks) [seconds]	0.03	0.345	0.00	0.505
		Timeouts	-0.06	0.792	0.03	0.330
		Errors of items which are incongruent	0.11	0.048	-0.05	0.756
	WORKING MEMORY					
N-back non verbal	NBN	Correct items	0.08	0.135	0.01	0.422
		Number of commission errors	0.08	0.121	0.00	0.497
		Number of errors	0.16	**0.009**	0.05	0.229
		Mean reaction tine of correct items [seconds]	0.00	0.505	-0.05	0.777
		Mean reaction time of errors [seconds]	-0.07	0.861	-0.01	0.574
Non-verbal learning test	NVLT	Sum of correct responses	0.16	**0.010**	-0.03	0.660
		Sum of false responses	0.18	**0.004**	0.07	0.163
		Sum of difference between correct minus false responses	0.24	**0.000**	0.21	**0.001**
		Process time	0.01	0.452	0.15	0.011
Corsi Block Tapping Test	CORSI	Block span	0.12	0.039	-0.05	0.771
		Correct items	0.15	0.012	-0.07	0.846
		False items	-0.05	0.772	-0.06	0.798
		Missed items	0.05	0.249	0.07	0.142
		Sequency errors	-0.04	0.728	0.05	0.220
	INHIBITION					
Stop Signal Task	INHIB	Reaction time [seconds]	0.15	0.010	0.12	0.037
		Mean stop signal delay [seconds]	0.04	0.276	0.14	0.018
		Stop signal reaction time [seconds]	0.09	0.103	0.09	0.099
		Number of commission errors	0.01	0.421	0.13	0.024
		Number of ommission errors	0.09	0.100	0.03	0.327
Simon Task	SIMON	Reaction time difference (reaction time incongruent - reaction time congruent items) [seconds]	0.08	0.126	-0.12	0.970
		Number of errors in compatible items	0.00	0.505	-0.03	0.689
		Number of errors in incompatible items	0.04	0.269	-0.02	0.618
Stroop Test	STROOP	Reading interference [seconds]	-0.04	0.719	-0.12	0.970
		Naming interference [seconds]	0.24	**0.000**	0.08	0.100
		Interference-difference [seconds]	0.13	0.024	-0.02	0.596
		Number of false reactions (reading-baseline)	0.12	0.036	-0.06	0.825
		Number of false reactions (naming-baseline)	0.09	0.086	-0.08	0.889
		Number of false reactions (reading-interference)	-0.03	0.690	-0.01	0.584
		Number of false reactions (naming-interference)	0.12	0.043	-0.08	0.897
		Process time	0.23	**0.000**	0.21	**0.001**
	ATTENTION / VIGILANCE					
Divided attention test	WAF-G	Number of missed items (unimodall visual)	0.06	0.190	-0.07	0.854
		Number of false alarm (unimodal visual)	0.04	0.291	0.03	0.353
		Mean reaction time (unimodal visual) [ms]	0.23	**0.000**	0.10	0.061
		Number of missed items (crossmodal visual/auditive)	0.21	**0.001**	0.16	**0.008**
		Number of false alarm (crossmodal visual/auditive)	0.20	**0.001**	0.12	0.032
		Mean reaction time (crossmodal) [ms]	0.27	**0.000**	0.08	0.128
Spatial attention test	WAF-R	Mean reaction time (unannounced items) [ms]	0.14	0.018	0.17	**0.004**
		Number of missed items (correct announced items)	-0.02	0.599	-0.03	0.653
		Mean reaction time (correct announced items) [ms]	0.21	**0.001**	-0.03	0.669
		Number of missed items (wrong announced items)	0.03	0.322	0.04	0.268
		Mean reaction time (wrong announced items) [ms]	0.08	0.120	0.14	0.018
		Mean reaction time (short SOA) [ms]	0.20	**0.002**	0.07	0.139
		Mean reaction time (long SOA) [ms]	0.19	**0.002**	0.08	0.111
		Number of errors	-0.07	0.849	-0.02	0.609
Mackworth Clock Test	MACK	Number of missed jumps	0.12	0.038	0.16	**0.009**
		Number of false alarms	0.03	0.352	-0.02	0.626

Supplementary Table S4: Correlations of all semantic verbal fluency features and significantly predicted EF scores (cognitive flexibility tests)

	SPM - Correct items		SPM - Process time		TMT - Process time partA		TMT - Process time partA		TMT - Difference B-A		WCST - Errors		WCST - Non-perseveration errors	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p
Correct words t1+t2+t3	0.100	0.143	-0.166	0.014	-0.111	0.103	-0.134	0.048	-0.103	0.130	-0.120	0.077	-0.157	0.020
Correct words t1	0.118	0.081	-0.069	0.312	-0.142	0.036	-0.092	0.176	-0.038	0.579	-0.086	0.207	-0.112	0.100
Correct words t2	0.014	0.834	-0.166	0.014	-0.008	0.904	-0.121	0.076	-0.138	0.042	-0.114	0.094	-0.158	0.020
Correct words t3	0.100	0.140	-0.197	0.003	-0.104	0.127	-0.112	0.099	-0.080	0.238	-0.084	0.215	-0.098	0.148
Repetition error t1	-0.052	0.442	0.039	0.571	-0.046	0.499	-0.034	0.614	-0.018	0.796	0.025	0.710	0.031	0.647
Repetition error t2	-0.044	0.522	-0.028	0.680	-0.064	0.346	-0.046	0.501	-0.022	0.746	0.027	0.688	0.017	0.800
Repetition error t3	-0.034	0.619	-0.086	0.207	0.003	0.964	-0.100	0.142	-0.119	0.080	-0.081	0.233	-0.104	0.124
Category error t3	-0.011	0.869	0.059	0.387	-0.021	0.761	-0.072	0.287	-0.075	0.271	-0.020	0.764	0.013	0.846
Latency mean t1	-0.158	0.020	0.020	0.765	0.129	0.058	0.041	0.548	-0.016	0.819	0.082	0.225	0.078	0.252
Latency mean t2	-0.022	0.746	0.110	0.105	-0.022	0.746	0.082	0.230	0.107	0.115	0.092	0.175	0.106	0.117
Latency mean t3	-0.106	0.120	0.127	0.061	0.103	0.129	0.075	0.269	0.037	0.584	0.067	0.323	0.064	0.349
Latencies 1st quarter t1	-0.130	0.055	0.101	0.138	0.129	0.057	0.130	0.054	0.089	0.188	0.211	0.002	0.165	0.014
Latencies 1st quarter t2	-0.043	0.531	0.118	0.082	-0.011	0.874	0.064	0.350	0.080	0.239	0.106	0.120	0.131	0.053
Latencies 1st quarter t3	0.038	0.581	0.150	0.027	0.115	0.092	-0.012	0.858	-0.071	0.296	0.046	0.499	0.060	0.379
Latencies 2nd quarter t1	-0.090	0.185	0.041	0.549	0.116	0.087	0.069	0.310	0.024	0.729	0.004	0.955	-0.032	0.641
Latencies 2nd quarter t2	0.009	0.898	0.121	0.074	0.001	0.986	0.078	0.249	0.092	0.178	0.053	0.438	0.093	0.171
Latencies 2nd quarter t3	-0.010	0.884	0.110	0.105	0.075	0.270	0.056	0.410	0.029	0.673	0.104	0.126	0.081	0.235
Latencies 3rd quarter t1	-0.048	0.478	0.047	0.495	0.065	0.342	-0.025	0.716	-0.061	0.369	-0.007	0.921	0.021	0.755
Latencies 3rd quarter t2	0.040	0.560	0.078	0.253	-0.010	0.883	0.038	0.572	0.050	0.461	0.003	0.960	-0.018	0.796
Latencies 3rd quarter t3	-0.058	0.394	0.063	0.358	0.008	0.905	-0.008	0.910	-0.013	0.848	-0.002	0.972	-0.027	0.688
Latencies 4th quarter t1	-0.139	0.041	-0.075	0.272	0.089	0.188	0.029	0.672	-0.010	0.879	0.048	0.478	0.107	0.114
Latencies 4th quarter t2	0.003	0.969	0.022	0.742	-0.067	0.323	-0.007	0.915	0.025	0.716	0.037	0.585	0.050	0.466
Latencies 4th quarter t3	-0.079	0.244	0.024	0.730	0.038	0.575	0.058	0.394	0.049	0.469	-0.038	0.577	-0.010	0.883
Latency difference t1	-0.129	0.057	-0.082	0.226	0.080	0.239	0.019	0.777	-0.017	0.803	0.033	0.631	0.095	0.161
Latency difference t2	0.008	0.911	0.009	0.898	-0.066	0.329	-0.015	0.828	0.016	0.819	0.025	0.714	0.035	0.612
Latency difference t3	-0.082	0.226	0.012	0.858	0.030	0.664	0.059	0.385	0.055	0.420	-0.042	0.541	-0.015	0.830
Sequential mean t1	0.028	0.686	0.025	0.711	0.077	0.260	-0.011	0.873	-0.051	0.456	0.089	0.192	0.082	0.226
Sequential mean t2	0.000	0.996	0.031	0.648	0.090	0.185	0.041	0.549	0.003	0.963	-0.077	0.259	-0.017	0.799
Sequential mean t3	0.010	0.884	0.039	0.565	-0.021	0.760	0.003	0.968	0.014	0.843	0.016	0.814	0.042	0.536
Cumulative mean t1	0.093	0.172	0.021	0.759	0.031	0.646	0.058	0.398	0.052	0.444	0.007	0.921	0.082	0.231
Cumulative mean t2	0.043	0.531	0.037	0.589	0.049	0.471	0.092	0.176	0.084	0.218	0.062	0.359	0.105	0.123
Cumulative mean t3	0.135	0.046	0.066	0.331	0.026	0.704	-0.048	0.479	-0.069	0.307	0.006	0.925	0.002	0.973
Sequential mean cat1 t3	0.100	0.141	0.074	0.276	-0.045	0.513	-0.055	0.416	-0.043	0.528	-0.071	0.299	-0.029	0.674
Sequential mean cat2 t3	0.062	0.361	-0.020	0.769	0.027	0.690	-0.076	0.264	-0.103	0.130	-0.094	0.165	-0.110	0.107
Switch coefficient	-0.017	0.807	-0.007	0.919	0.027	0.688	0.049	0.474	0.044	0.520	0.061	0.366	0.094	0.165
Sequential mean t1 DIS	0.075	0.272	0.047	0.495	0.110	0.106	0.161	0.017	0.135	0.047	0.018	0.796	0.022	0.743
Sequential mean t2 DIS	0.020	0.768	0.019	0.778	0.043	0.524	-0.010	0.886	-0.033	0.628	0.001	0.988	-0.030	0.655
Sequential mean t3 DIS	-0.075	0.267	0.035	0.608	-0.008	0.905	-0.039	0.571	-0.041	0.544	0.013	0.844	0.011	0.873
Sequential mean cat1 t3 DIS	-0.046	0.495	0.025	0.714	-0.004	0.955	0.088	0.195	0.106	0.120	0.016	0.810	0.038	0.576
Sequential mean cat2 t3 DIS	0.038	0.580	0.005	0.944	0.035	0.607	-0.076	0.265	-0.107	0.117	-0.063	0.357	-0.025	0.717
Cumulative mean t1 DIS	-0.031	0.652	-0.056	0.408	0.108	0.111	0.139	0.040	0.110	0.106	0.038	0.576	0.065	0.339
Cumulative mean t2 DIS	0.001	0.991	0.036	0.598	0.033	0.632	0.039	0.566	0.030	0.662	0.010	0.878	0.024	0.722
Cumulative mean t3 DIS	0.087	0.199	0.093	0.170	0.004	0.955	-0.021	0.758	-0.027	0.696	0.048	0.480	0.071	0.300

t1=VF test (animals); t2= VF test (jobs); t3= Switching VF test (sports/fruits); cat1=1st category in VF switching task (sports); cat2=2nd category in VF switching task (fruits); DIS=semantic system *DISCO*

Supplementary Table S5: Correlations of all semantic verbal fluency features and significantly predicted EF scores (working memory and inhibition tests)

	NBN - Errors		NVLT - Correct		NVLT - Errors		NVLT - Difference correct-false		STROOP - Naming interference		STROOP - Process time	
	r	p	r	p	r	p	r	p	r	p	r	p
Correct words t1+t2+t3	0.039	0.563	0.051	0.456	-0.001	0.992	0.033	0.631	-0.221	0.001	-0.194	0.004
Correct words t1	0.008	0.910	0.024	0.727	-0.059	0.388	0.074	0.275	-0.152	0.025	-0.160	0.018
Correct words t2	0.080	0.242	0.026	0.701	0.073	0.284	-0.057	0.403	-0.212	0.002	-0.101	0.136
Correct words t3	0.004	0.959	0.092	0.176	0.000	1.000	0.058	0.394	-0.163	0.016	-0.218	0.001
Repetition error t1	0.107	0.114	0.009	0.899	0.135	0.046	-0.131	0.054	-0.039	0.564	-0.011	0.877
Repetition error t2	0.054	0.431	-0.074	0.277	0.161	0.017	-0.209	0.002	-0.076	0.265	-0.040	0.552
Repetition error t3	-0.008	0.907	-0.089	0.190	-0.032	0.641	-0.024	0.721	0.024	0.725	0.003	0.966
Category error t3	-0.007	0.919	0.022	0.741	0.047	0.490	-0.033	0.626	0.025	0.717	-0.060	0.380
Latency mean t1	-0.030	0.656	-0.072	0.293	0.083	0.221	-0.129	0.057	0.074	0.280	0.076	0.267
Latency mean t2	-0.062	0.360	-0.014	0.840	-0.035	0.605	0.027	0.694	0.155	0.022	0.031	0.649
Latency mean t3	0.008	0.907	-0.090	0.187	0.037	0.586	-0.094	0.167	0.159	0.019	0.179	0.008
Latencies 1st quarter t1	-0.027	0.697	-0.048	0.479	-0.027	0.688	-0.003	0.967	0.185	0.006	0.144	0.033
Latencies 1st quarter t2	-0.032	0.634	0.046	0.500	-0.052	0.445	0.081	0.231	0.165	0.015	0.110	0.104
Latencies 1st quarter t3	-0.008	0.906	-0.052	0.445	0.035	0.608	-0.068	0.317	0.119	0.081	0.124	0.069
Latencies 2nd quarter t1	0.004	0.954	0.044	0.515	0.098	0.150	-0.071	0.300	-0.012	0.855	0.017	0.804
Latencies 2nd quarter t2	-0.074	0.277	-0.018	0.790	-0.092	0.176	0.081	0.232	0.107	0.116	0.027	0.692
Latencies 2nd quarter t3	0.113	0.095	-0.024	0.723	0.064	0.344	-0.080	0.239	0.172	0.011	0.194	0.004
Latencies 3rd quarter t1	-0.033	0.624	-0.079	0.244	0.047	0.494	-0.097	0.153	0.033	0.632	0.064	0.345
Latencies 3rd quarter t2	0.040	0.557	-0.032	0.640	-0.004	0.950	-0.016	0.817	0.021	0.762	-0.035	0.605
Latencies 3rd quarter t3	-0.085	0.210	0.003	0.963	-0.022	0.751	0.024	0.727	0.068	0.318	0.011	0.871
Latencies 4th quarter t1	-0.056	0.408	-0.083	0.222	0.065	0.341	-0.118	0.083	0.048	0.484	0.055	0.421
Latencies 4th quarter t2	-0.016	0.816	-0.010	0.882	0.031	0.652	-0.037	0.583	0.104	0.127	-0.017	0.804
Latencies 4th quarter t3	0.052	0.442	-0.052	0.441	-0.025	0.715	-0.008	0.906	-0.063	0.353	-0.003	0.960
Latency difference t1	-0.054	0.423	-0.080	0.242	0.067	0.325	-0.118	0.083	0.034	0.617	0.044	0.516
Latency difference t2	-0.012	0.859	-0.016	0.819	0.037	0.587	-0.047	0.489	0.085	0.212	-0.030	0.661
Latency difference t3	0.053	0.435	-0.049	0.475	-0.028	0.685	-0.003	0.966	-0.072	0.287	-0.013	0.850
Sequential mean t1	-0.192	0.005	-0.026	0.699	0.017	0.806	-0.033	0.623	-0.016	0.814	-0.004	0.959
Sequential mean t2	0.045	0.505	-0.104	0.127	0.007	0.921	-0.072	0.288	-0.181	0.008	-0.030	0.660
Sequential mean t3	-0.033	0.629	-0.017	0.799	-0.037	0.592	0.026	0.705	-0.027	0.688	-0.029	0.668
Cumulative mean t1	-0.014	0.842	0.015	0.825	-0.052	0.441	0.062	0.359	-0.175	0.010	-0.170	0.012
Cumulative mean t2	0.053	0.433	-0.039	0.568	0.014	0.832	-0.039	0.565	-0.184	0.006	-0.113	0.096
Cumulative mean t3	0.001	0.992	0.037	0.583	-0.041	0.547	0.065	0.340	-0.181	0.007	-0.204	0.002
Sequential mean cat1 t3	-0.114	0.094	0.037	0.582	0.018	0.792	0.006	0.935	-0.120	0.078	-0.127	0.061
Sequential mean cat2 t3	-0.015	0.829	0.014	0.838	-0.031	0.652	0.040	0.560	0.020	0.765	-0.056	0.412
Switch coefficient	-0.048	0.483	0.038	0.576	0.001	0.990	0.023	0.734	0.104	0.126	0.006	0.924
Sequential mean t1 DIS	-0.087	0.201	0.004	0.954	0.017	0.802	-0.015	0.829	-0.064	0.349	-0.015	0.821
Sequential mean t2 DIS	0.083	0.223	0.082	0.226	0.095	0.161	-0.044	0.517	-0.169	0.012	0.002	0.980
Sequential mean t3 DIS	0.033	0.631	0.013	0.847	-0.052	0.447	0.061	0.374	-0.010	0.880	-0.009	0.896
Sequential mean cat1 t3 DIS	-0.081	0.234	0.097	0.153	0.121	0.075	-0.060	0.374	-0.147	0.030	-0.065	0.341
Sequential mean cat2 t3 DIS	-0.128	0.060	0.048	0.478	-0.083	0.220	0.115	0.091	0.015	0.825	-0.021	0.763
Cumulative mean t1 DIS	-0.004	0.953	-0.059	0.384	0.045	0.505	-0.083	0.221	0.014	0.840	-0.001	0.992
Cumulative mean t2 DIS	-0.036	0.597	0.075	0.270	-0.074	0.279	0.122	0.073	-0.073	0.282	-0.008	0.902
Cumulative mean t3 DIS	-0.123	0.069	-0.016	0.819	-0.082	0.227	0.073	0.284	-0.105	0.121	-0.058	0.393

t1=VF test (animals); t2= VF test (jobs); t3= Switching VF test (sports/fruits); cat1=1st category in VF switching task (sports); cat2=2nd category in VF switching task (fruits); DIS=semantic system *DISCO*

Supplementary Table S6: Correlations of all semantic verbal fluency features and significantly predicted EF scores (attention tests)

	WAF-G - rt unimodal		WAF-G - Missed items crossmodal		WAF-G - False alarm crossmodal		WAF- rt crossmodal		WAF-R - rt correct announced		WAF-R - rt short SOA		WAF-R - rt long SOA	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p
Correct words t1+t2+t3	-0.019	0.783	-0.124	0.068	-0.067	0.326	-0.073	0.286	0.007	0.922	-0.067	0.321	-0.035	0.602
Correct words t1	-0.026	0.705	-0.180	0.008	-0.166	0.014	-0.095	0.163	-0.021	0.759	-0.058	0.390	-0.047	0.490
Correct words t2	0.047	0.490	0.002	0.972	0.075	0.271	0.016	0.819	0.024	0.719	-0.046	0.496	-0.005	0.947
Correct words t3	-0.086	0.204	-0.092	0.175	-0.031	0.654	-0.096	0.158	0.026	0.704	-0.054	0.425	-0.027	0.687
Repetition error t1	-0.035	0.604	0.043	0.528	0.019	0.776	-0.052	0.447	-0.092	0.174	-0.089	0.190	-0.095	0.164
Repetition error t2	0.045	0.512	0.027	0.695	0.120	0.077	-0.027	0.697	-0.006	0.926	-0.037	0.590	0.008	0.905
Repetition error t3	-0.118	0.083	-0.029	0.674	-0.013	0.852	-0.093	0.171	0.011	0.875	-0.004	0.952	0.031	0.652
Category error t3	-0.026	0.701	0.025	0.718	0.073	0.286	-0.034	0.620	-0.041	0.545	-0.040	0.560	-0.041	0.543
Latency mean t1	0.019	0.783	0.121	0.075	0.063	0.352	0.056	0.408	-0.069	0.312	-0.028	0.686	-0.038	0.581
Latency mean t2	-0.057	0.402	0.005	0.942	-0.081	0.235	-0.045	0.505	-0.009	0.899	0.055	0.420	0.003	0.969
Latency mean t3	0.070	0.302	0.097	0.155	0.041	0.544	0.096	0.157	-0.019	0.775	0.035	0.603	0.007	0.919
Latencies 1st quarter t1	0.057	0.402	0.061	0.372	0.008	0.901	-0.023	0.730	-0.015	0.825	0.012	0.859	0.028	0.677
Latencies 1st quarter t2	-0.043	0.530	0.047	0.491	-0.042	0.534	-0.002	0.980	0.013	0.850	0.044	0.515	0.037	0.590
Latencies 1st quarter t3	0.091	0.182	0.001	0.991	0.030	0.656	0.127	0.061	-0.108	0.112	-0.027	0.695	-0.069	0.311
Latencies 2nd quarter t1	0.063	0.356	0.038	0.574	0.015	0.830	0.141	0.038	-0.029	0.672	0.013	0.850	-0.007	0.916
Latencies 2nd quarter t2	-0.031	0.645	-0.097	0.152	-0.065	0.340	0.020	0.771	-0.062	0.365	0.009	0.893	-0.043	0.529
Latencies 2nd quarter t3	0.022	0.752	0.084	0.215	0.049	0.476	0.046	0.499	0.043	0.531	0.088	0.194	0.053	0.440
Latencies 3rd quarter t1	0.019	0.784	0.098	0.151	0.079	0.246	0.028	0.683	-0.045	0.507	-0.013	0.843	-0.039	0.571
Latencies 3rd quarter t2	-0.056	0.409	0.033	0.628	-0.036	0.594	-0.085	0.211	0.024	0.729	0.038	0.572	-0.005	0.939
Latencies 3rd quarter t3	-0.010	0.883	0.003	0.959	-0.045	0.506	-0.003	0.963	-0.042	0.536	-0.010	0.882	-0.008	0.903
Latencies 4th quarter t1	-0.050	0.459	0.163	0.016	0.122	0.071	0.086	0.209	0.028	0.677	0.020	0.764	0.027	0.688
Latencies 4th quarter t2	-0.007	0.922	0.079	0.245	0.045	0.512	-0.097	0.155	0.058	0.396	0.066	0.334	0.015	0.823
Latencies 4th quarter t3	-0.032	0.636	0.062	0.364	-0.008	0.903	-0.052	0.446	0.012	0.860	-0.005	0.940	0.007	0.923
Latency difference t1	-0.055	0.421	0.158	0.019	0.122	0.072	0.087	0.199	0.030	0.665	0.020	0.774	0.025	0.710
Latency difference t2	-0.002	0.980	0.074	0.277	0.050	0.464	-0.097	0.154	0.057	0.406	0.061	0.371	0.011	0.872
Latency difference t3	-0.039	0.565	0.062	0.362	-0.011	0.876	-0.062	0.364	0.020	0.766	-0.003	0.964	0.012	0.861
Sequential mean t1	0.048	0.483	-0.008	0.903	-0.121	0.074	-0.004	0.951	0.007	0.913	0.050	0.459	0.037	0.588
Sequential mean t2	0.041	0.545	0.010	0.886	0.018	0.795	-0.057	0.406	-0.002	0.982	-0.004	0.952	0.000	0.995
Sequential mean t3	-0.103	0.130	-0.039	0.564	-0.107	0.116	-0.077	0.256	-0.048	0.482	-0.058	0.396	-0.061	0.371
Cumulative mean t1	-0.019	0.776	-0.094	0.169	-0.176	0.009	-0.036	0.593	0.059	0.383	0.055	0.415	0.059	0.387
Cumulative mean t2	0.032	0.639	0.078	0.249	0.038	0.579	-0.013	0.845	0.002	0.976	0.010	0.883	-0.013	0.848
Cumulative mean t3	-0.061	0.371	-0.030	0.662	-0.089	0.189	-0.049	0.475	-0.007	0.913	-0.039	0.570	-0.062	0.362
Sequential mean cat1 t3	-0.066	0.331	-0.114	0.094	-0.065	0.339	-0.048	0.479	-0.063	0.355	-0.057	0.405	-0.083	0.222
Sequential mean cat2 t3	-0.167	0.014	-0.127	0.060	-0.025	0.714	-0.108	0.110	0.013	0.853	-0.029	0.665	-0.046	0.496
Switch coefficient	-0.074	0.279	0.058	0.397	0.053	0.436	-0.014	0.839	0.021	0.760	0.027	0.695	0.015	0.822
Sequential mean t1 DIS	0.012	0.865	0.007	0.921	-0.042	0.536	-0.006	0.933	0.035	0.607	0.033	0.628	0.021	0.759
Sequential mean t2 DIS	0.080	0.237	-0.002	0.982	0.020	0.766	0.111	0.101	0.036	0.593	0.044	0.522	0.068	0.321
Sequential mean t3 DIS	-0.021	0.756	0.016	0.815	-0.046	0.498	0.018	0.792	0.149	0.028	0.121	0.076	0.101	0.138
Sequential mean cat1 t3 DIS	-0.019	0.783	-0.030	0.656	-0.025	0.715	-0.065	0.343	-0.064	0.345	-0.067	0.326	-0.108	0.112
Sequential mean cat2 t3 DIS	-0.073	0.285	-0.052	0.443	-0.044	0.515	0.058	0.393	-0.035	0.609	-0.059	0.384	-0.057	0.404
Cumulative mean t1 DIS	-0.007	0.919	0.125	0.065	0.173	0.010	0.048	0.482	0.072	0.287	0.049	0.475	0.076	0.266
Cumulative mean t2 DIS	0.048	0.480	0.007	0.916	-0.047	0.494	0.063	0.357	0.101	0.135	0.105	0.123	0.102	0.135
Cumulative mean t3 DIS	-0.073	0.280	-0.066	0.332	-0.102	0.133	-0.007	0.922	-0.039	0.563	-0.059	0.385	-0.073	0.286

t1=VF test (animals); t2= VF test (jobs); t3= Switching VF test (sports/fruits); cat1=1st category in VF switching task (sports); cat2=2nd category in VF switching task (fruits); DIS=semantic system *DISCO*