Original Research Article

On Which Abilities Are Category Fluency and Letter Fluency Grounded? A Confirmatory Factor Analysis of 53 Alzheimer's Dementia Patients

Ilaria Bizzozero[b] Stefania Scotti[a, b] Francesca Clerici[c] Simone Pomati[c]
Marcella Laiacona[d] Erminio Capitani[a, b]

[a]Neurology Unit, Health Sciences Department, [b]Neurology Unit, S. Paolo Hospital, and [c]Centre for Research and Treatment of Cognitive Dysfunctions, Neurology Unit, L. Sacco Hospital, University of Milan, Milan, and [d]Units of Aphasia and Neurology, Fondazione Salvatore Maugeri, IRCCS, Istituto Scientifico di Veruno, Novara, Italy

Key Words

Language · Dementia · Verbal fluency · Semantics · Attention

Abstract

Background/Aims: In Alzheimer's dementia (AD), letter fluency is less impaired than category fluency. To check whether category fluency and letter fluency depend differently on semantics and attention, 53 mild AD patients were given animal and letter fluency tasks, two semantic tests (the Verbal Semantic Questionnaire and the BORB Association Match test), and two attentional tests (the Stroop Colour-Word Interference test and the Digit Cancellation test). ***Methods:*** We conducted a LISREL confirmatory factor analysis to check the extent to which category fluency and letter fluency tasks were related to semantics and attention, viewed as latent variables. ***Results:*** Both types of fluency tasks were related to the latent variable Semantics but not to the latent variable Attention. ***Conclusions:*** Our findings warn against interpreting the disproportionate impairment of AD patients on category and letter fluency as a contrast between semantics and attention.

Copyright © 2013 S. Karger AG, Basel

Erminio Capitani, MD
Neurology Unit, S. Paolo Hospital, University of Milan
via Di Rudinì 8
IT–20142 Milan (Italy)
E-Mail erminio.capitani @ unimi.it

KARGER

180

Dement Geriatr Cogn Disord Extra 2013;3:179–191

DOI: 10.1159/000351418 | © 2013 S. Karger AG, Basel
www.karger.com/dee

Bizzozero et al.: On Which Abilities Are Category Fluency and Letter Fluency
Grounded? A Confirmatory Factor Analysis of 53 Alzheimer's Dementia Patients

Introduction

Verbal fluency tasks are widely used for the assessment of cognitive impairment. Patients are often examined on both category fluency and letter fluency [1, 2]. The discrepant pattern observed in different pathological groups seems consistent with the general claim that the deficits in category fluency and letter fluency have distinct roots: Alzheimer's dementia (AD) patients are relatively more impaired on category fluency, whereas traumatic brain injury patients show the opposite pattern and commit more errors on letter fluency [1].

Given the experimental and clinical relevance of the two fluency variants, it is worth probing in greater depth their basic neuropsychological components. Category fluency seems largely based on an 'automatic' search through the lexical-semantic system, devoid of attention load. On the contrary, to perform the letter fluency task, it is necessary to implement a more effortful lexical search based on the first phoneme of the target. This explanation would be in accordance with the notion that AD patients, who are more impaired on category fluency, present a severe deficit in lexical and semantic competence. In contrast, traumatic brain injury patients, who are more impaired on letter fluency, have considerable attention deficits.

However, a rigid dichotomy between semantics and attention is probably an oversimplification, and on all tasks requiring any type of name retrieval, semantics, lexicon and control functions are likely to interact. Reverberi et al. [3] examined a series of frontal patients on category fluency and concluded that their deficit had definite 'strategical' or 'attentional' components changing according to the site of lesion (dorso-lateral vs. mesial). The general relevance of attentional and control functions has been widely advocated in the study of AD patients. Faust et al. [4] claimed that picture-naming errors of AD patients are due to a breakdown of the access to phonological representations of object names as a consequence of the reduced inhibitory control over other highly active alternatives. This interpretation is consistent with the suggestion that semantic knowledge is not directly impaired in AD, but rather not readily accessible or usable if the cognitive processes involved in a task make heavy demands on the attentional capacities [5, 6].

Besides suggesting that attentional/control skills might be relevant for category fluency, the literature also indicates that semantic and lexical abilities can be crucial for letter fluency. Schwartz et al. [7] conducted a fluency experiment on normal participants, and found that during the letter fluency task, there was a striking categorical separation between animate and inanimate entities. They concluded that even on letter fluency, the semantic facilitation was pervasive, and challenged the traditional view of letter fluency as a purely phonemically based task.

Considering this debate, the reason why AD patients perform better on letter fluency than on category fluency might depend not only on a relatively greater impairment of semantic and lexical abilities, but also on a relatively greater deficit in attentional control. Besides the studies by Nebes et al. [5, 6], Chenery et al. [8] found that in the early stage of AD, decline in performance was due to changes in attentional control. Therefore, the attentional deficit hypothesis would require a first check in patients who are still in the early stages of AD.

Summing up, attentional/control skills and lexical and semantic ability still need to be disentangled in the attempt to better understand the performance of AD patients on verbal fluency. To this end, we ran an experiment on a fresh sample of mild-to-moderate AD patients using confirmatory factor analysis [9]. Our question was whether, among AD patients, the semantic and attention loads of category or letter fluency are disproportionate.

Let us first suppose that a stringent statistical analysis confirms that category fluency mainly taps semantics, whereas letter fluency taps attention. In this case, the disproportionate impairment of category fluency would be explained if mild-to-moderate AD patients were more severely impaired on semantics than on attention.

Bizzozero et al.: On Which Abilities Are Category Fluency and Letter Fluency Grounded? A Confirmatory Factor Analysis of 53 Alzheimer's Dementia Patients

Alternatively, mild AD patients might suffer from a prevailing attentional deficit. The automatic, uninhibited activation of words phonologically similar to the last uttered name [4], viewed as one facet of general attention impairment, could impair the performance on category fluency and paradoxically improve the letter fluency task. This explanation would be empirically supported if (i) category fluency was substantially and directly dependent on attention, and (ii) mild-to-moderate AD patients were more severely impaired on attention tests than on semantic tests.

As a direct comparison of the importance of semantics and attention across the two fluency types is missing in the literature, this analysis is crucial for a correct interpretation of clinical data.

Methods

Participants

Fifty-three AD patients participated in this experiment. This research was completed in accordance with the Helsinki declaration and with the ethical rules of the San Paolo Hospital and the L. Sacco Hospital of the University of Milan. The participants were outpatients of two dementia units (Neurology Units at the S. Paolo Hospital and the L. Sacco Hospital), and they received a diagnosis of AD based on the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) and on NINCS-ADRDA criteria. Until the completion of testing, the participants did not begin any therapy for the treatment of cognitive decline. We included only patients affected by a cognitive decline of mild-to-moderate severity, as quantified by a MODA scale score [10] of at least 67.5/100 (corresponding to a MMSE score of 18.0 [11]); the upper severity limit was a MODA scale score of 89.0/100, i.e. the pathology/normality boundary. There was no overlap between the AD sample investigated here and that studied by Capitani et al. [1].

Examination Battery

Participants underwent the following tests.

(i) Category Fluency with semantic cue (animals) of 1-min duration. The score consisted of the number of animal names produced. For a procedure description and reference norms, see Capitani et al. [1]. Age- and education-adjusted scores were considered pathological when they were lower than 12.5. Henceforth, 'Category Fluency' (with capitals) will be used with reference to the test actually used in this study.

(ii) Letter Fluency with phonological cue (the three initial phonemes were f, p, and l). The score corresponded to the mean number of words uttered on the three trials corresponding to each initial letter. For a procedure description and reference norms, see Capitani et al. [1]. Age- and education-adjusted scores were considered pathological when they were lower than 7.1. Henceforth, 'Letter Fluency' (with capitals) will be used with reference to the test actually used in this study.

(iii) The Verbal Semantic Questionnaire [12]. The patients were given 10 animal names and were asked to respond to six questions for each of these stimuli (with a hit score range of 0–60). The six questions referred to different levels of knowledge: (a) general superordinate, (b) same category superordinate, (c) subordinate: perceptual attribute, (d) subordinate: comparison of perceptual attributes, (e) subordinate: associative functional aspects, and (f) subordinate: associative contextual aspects. The subjects were asked to select the correct response among two or three orally presented alternatives. One point was given for each correct answer. Reference norms have been calculated on a sample of 60 elderly with a low-to-intermediate education level [12]: a score of 51/60 or less was considered pathological.

KARGER

182

Dement Geriatr Cogn Disord Extra 2013;3:179–191

DOI: 10.1159/000351418 | © 2013 S. Karger AG, Basel
www.karger.com/dee

Bizzozero et al.: On Which Abilities Are Category Fluency and Letter Fluency
Grounded? A Confirmatory Factor Analysis of 53 Alzheimer's Dementia Patients

(iv) The BORB Association Match test [13]. From the whole series of 30 trials, we considered only the six trials that tested animal knowledge. The possible score range was 0–6. For this version of the battery, as norms were not available, we collected an ad hoc sample of 50 normal subjects matched to the AD group for the relevant demographic variables (mean age = 72.3, SD = 6.4; mean education years = 8.3, SD = 6.3; males/females = 0.9); on this basis, scores were considered normal if at least five hits were registered. The same sample was used to calculate a normality threshold for the Stroop Colour-Word Interference (CWI) test (see below).

(v) The Stroop CWI test [14]. This version of the test consisted of three parts. In the first part, the participant was asked to read 30 colour names printed with black ink, in the second part, the participant had to name 30 colour patches ('basic condition'), and in the final section ('interfered condition'), the ink colour of the colour names (red, blue, or green) printed with a colour different from the word content had to be named. In each section, the 30 stimuli were presented on three white paper sheets of 21 × 29 cm, each containing 10 stimuli arranged vertically. There was no time limit. We calculated a unique measure of the CWI obtained by considering (i) the time needed to complete the naming of all 30 coloured patches of the basic condition, and (ii) the time necessary for naming the ink colour of all 30 colour names of the interfered condition. Starting from these two time measures and from the two respective numbers of hits, we divided for each section the completing time by the number of hits, and finally, subtracted the time/hit ratio of the basic condition from that of the interfered condition. Thus, the score indicates the additive time needed for the correct naming of a single stimulus in the interfered condition with reference to the basic condition: higher scores correspond to higher interference and worse performance. To avoid statistical problems arising from the presence of two outlying observations (scores of 30 and 31), we replaced these interference scores with the third-last score (a score of 15) that immediately preceded the two outlying observations in the series of ranked increasing scores. Normality values were collected ad hoc from the sample of 50 normal controls described above. Based on the control group performance, scores less than or equal to 2 s were considered normal.

(vi) The Digit Cancellation test [15]. In this study, patients were given only the first section, with a time limit of 45 s. Each subject was presented with a matrix of 13 rows containing 10 numbers from 0 to 9, and was asked to scan the matrix from the first to the last row and to cross all the number '5' printed in the matrix. The first three rows were considered as 'run in' stimuli. The hits range was 0–10. Norms were those currently used in our laboratory; they were calculated from a sample of 252 normal subjects (mean age = 51.5, SD = 15.5; mean education years = 10.5, SD = 4.9; males = 50%) available from our records, after adjustment for education (age was not influential) through the setting of one-sided non-parametric tolerance limits (E. Capitani, unpubl. data; for a methodological introduction to non-parametric tolerance limits see Capitani and Laiacona [16]). In order to be consistent with the other attention measure (the Stroop CWI test), the score consisted of the number of missed targets. For this test, adjusted scores higher than 4 were considered pathological, whereas scores between 2.7 and 4 were considered borderline. None of the controls produced false alarms.

The assembly of our experimental battery was guided by the aim to represent different facets of the broad constructs of semantics and attention: for semantics, we chose an overtly verbal test (the Verbal Semantic Questionnaire) and a semantic test (the BORB Association Match test) based solely on pictorial stimuli. For attention, we chose a classical test sensitive to interference (the Stroop CWI test) and a more basic test of selective attention (the Digit Cancellation test). For both cognitive dimensions, we strove to assemble a battery that could be effectively given to AD patients, and necessarily limited the number of included tests to the minimum in order to avoid dropouts. The scores of four of our experimental measures (the two fluency tasks, the BORB Association Match test, and the Verbal Semantic Questionnaire)

Bizzozero et al.: On Which Abilities Are Category Fluency and Letter Fluency
Grounded? A Confirmatory Factor Analysis of 53 Alzheimer's Dementia Patients

were proficiency scores. The two remaining measures (the Stroop CWI and the Digit Cancellation tests) were 'error measures', and higher values of these measures correspond to poorer performance. It would have been advisable to score all the six observed measures listed above either as 'proficiency scores' or as 'error scores'. However, this was not possible because fluency tasks and the Stroop CWI test, in principle, are open scale measures, and for the former case, a high number of words means a good performance, whereas in the latter case, long interference times mean poor attention.

General Approach and Statistical Methodology

Our approach was based on confirmatory factor analysis [9]. It consists of the construction of a model of interrelations between a set of variables: one part is directly observed during the experiment, while another part is not directly observed and therefore considered 'latent'. Each latent variable may be defined as the construct underlying a subset of observed variables, and the composition of this subset is decided by the examiner and is specific to a given model. A correlation matrix calculated on the observed variables is used as an input. The analysis of each model yields the estimation of the parameters expressing the relationship between each latent variable and the observed variables used for its definition, and the procedure also quantifies the relationship between the latent variables. For a recent neuropsychological application of this methodology see, for example, Barbey et al. [17].

For each of the models, the statistical analysis provides a goodness-of-fit measure that follows the χ^2 distribution. This statistic expresses the extent to which the model at issue fails to explain the relationship between the variables; consequently, the best explanation is provided by the model associated with the lowest χ^2 (interpreted based on its degree of freedom, d.f.). Models that differ as regards the inclusion/exclusion of a link between variables can be directly compared by calculating the difference between their respective goodness-of-fit statistics.

In our case, two latent variables, one tapping semantics and one tapping attention, were defined by the performance of a group of AD patients on two overtly semantic tasks *plus* the fluency tasks on the one hand, and by two overtly attentional tasks *plus* the fluency tasks on the other. These latent variables will be indicated with the terms Semantics and Attention. We started from a 'neutral' model where Category Fluency required both semantic and attentional competence, and Letter Fluency required both attentional and semantic competence. After assessing the overall goodness-of-fit of this model, we removed, one at the time, the four parameters expressing the relationship between each type of fluency and each latent variable. The purpose of these subtractions was to check whether the overall goodness-of-fit significantly worsened if a given link was removed from the model. For example, if category fluency crucially depends on attention, the goodness-of-fit would visibly decline when removing the link between Category Fluency and the latent variable Attention, and so on for each of the four parameters that link each fluency variant to each latent variable.

Results

Descriptive Statistics

Mean age of the 53 AD patients was 77.4 years (SD = 5.9, range 64–87). Mean education was 7.6 years (SD = 3.8, range 3–17). Seventeen patients were males, 36 were females. The mean adjusted MODA score was 77.8/100 (SD = 6.0, range 67.5–89.0), and the pathology threshold, adjusted for age and education, was 85.5 or less [10]; scores between 85.5 and 89 were rated as borderline. In our sample, 7 AD patients scored in the borderline range (13.2%), and the remainders (46 patients, corresponding to 86.8%) were pathological.

Bizzozero et al.: On Which Abilities Are Category Fluency and Letter Fluency
Grounded? A Confirmatory Factor Analysis of 53 Alzheimer's Dementia Patients

Table 1. Descriptive statistics

	Category Fluency	Letter Fluency (mean of 3 letters)	BORB Association Match test	Verbal Semantic Questionnaire	Stroop CWI test	Digit Cancellation test
Mean score ± SD	8.0±2.6	6.3±2.4	5.0±1.0	48.8±6.0	3.6±3.6	0.9±1.2
Theoretical range	0/n.u.l.	0/n.u.l.	0–6 c.r.	0–60 c.r.	interference: 0–15 s	0–10 missed targets
Observed range	2–13	1.3–10.7	3–6 c.r.	29–58 c.r.	interference: 0.2–15 s	error score: 0–5
Pathology threshold	<12.5	<7.1	<5	<52	>2 s	>4.0
Pathological scores	30/53 (56.7%)	22/53 (41.5%)	13/53 (24.5%)	34/53 (64.1%)	32/53 (60.4%)	2/53 (3.8%)

For each test, the theoretically poorest and best scores are indicated. On the Stroop CWI test and the Digit Cancellation test, higher scores correspond to poorer performance. n.u.l. = No upper limit; c.r. = correct responses.

Table 1 shows the mean of the experimental variables and the percentage of subjects presenting a pathological performance on each test according to the available norms. Some tests were more sensitive than others to the presence of AD, but this gradient did not conform to the general distinction between semantic tests and tests tapping attention/control functions. Category Fluency was more frequently impaired (56.7%) than Letter Fluency (41.5%). On the McNemar test, there were significantly more AD patients showing errors on Category Fluency only (n = 11) than on Letter Fluency only (n = 3), with a two-sided p value of 0.033. Taking the available norms as a reference [1], we calculated the mean adjusted fluency scores of the AD group and the z scores corresponding to the discrepancy between the control means and the AD group means. The mean z score of Category Fluency was –1.601, and that of Letter Fluency was –1.176. Accordingly, on Category Fluency, the patients were not only more frequently under the threshold, but were also more severely impaired on average.

Apart from the two fluency tasks, a formal statistically controlled comparison between the impairment rates in different tests was not always advisable because, for some of the tests, norms derived from less extended samples matched ad hoc to our AD group, while for others they derived from previously available larger samples adjusted for the demographic variables. Although thresholds were set following different methodologies, their inspection is still informative. The Verbal Semantic Questionnaire and the Stroop CWI test were clearly more impaired than the BORB Association Match and the Digit Cancellation tests.

Correlation between the Experimental Variables

Before calculating the correlation between the experimental variables, the original score of the Stroop CWI test was submitted to a logarithmic transformation in order to normalize the score distribution. Table 2 shows the correlation matrix between the six variables directly observed in this experiment. The optimal way to quantify the correlation between two variables depends on the scale on which each score is set: when both scales are continuous, the Pearson coefficient is appropriate; when one scale is continuous and the other ordinal, polyserial correlation should be used; and finally, when both scales are ordinal, polychoric correlation is appropriate [9]. When the range of the score observed with a given measure is narrow, it is conservative to consider this measure as an ordinal score. In our study, the BORB Association Match and the Digit Cancellation tests were treated as ordinal scores throughout the statistical analysis.

185

EXTRA
Dementia
and Geriatric
Cognitive Disorders

Dement Geriatr Cogn Disord Extra 2013;3:179–191

DOI: 10.1159/000351418 | © 2013 S. Karger AG, Basel
www.karger.com/dee

Bizzozero et al.: On Which Abilities Are Category Fluency and Letter Fluency
Grounded? A Confirmatory Factor Analysis of 53 Alzheimer's Dementia Patients

Table 2. Correlation matrix between the six experimental variables of this study

	Category Fluency	Letter Fluency	BORB Association Match test[1]	Verbal Semantic Questionnaire	Stroop CWI test (logarithmic)
Letter Fluency	PE = +0.322				
BORB Association Match test[1]	PS = +0.456	PS = +0.385			
Verbal Semantic Questionnaire	PE = +0.240	PE = +0.418	PS = +0.526		
Stroop CWI test (logarithmic)	PE = –0.279	PE = –0.385	PS = –0.228	PE = –0.326	
Digit Cancellation test[1]	PS = –0.128	PS = –0.255	PC = –0.216	PS = –0.236	PS = +0.563

Scores of the semantic tests and Letter Fluency are directly proportional to proficiency. The Stroop CWI test and the Digit Cancellation test are inversely proportional to proficiency. The significance level of the correlation coefficients is not indicated because the significance estimation is not strictly comparable for Pearson's correlation, polyserial and polychoric correlations. As a general reference, for a sample size of n = 53, a Pearson's correlation of 0.228 corresponds to $p = 0.10$, a correlation of 0.271 to $p = 0.05$ and a correlation of 0.351 to $p = 0.01$. PC = Polychoric correlation coefficient; PE = Pearson's product-moment correlation; PS = polyserial correlation coefficient.

[1] An ordinal measurement scale was assumed for these tests.

LISREL Analysis

The resulting correlation matrix was used as the input for a LISREL analysis [9]. Table 3 shows the LISREL analysis of the starting model (with eight links) and of the four different models obtained by removing, one at a time, the links that connected a given fluency type with a given latent variable. In this set of analyses, the parameter estimation was always characterized by good convergence of the iterative procedure, and negative error variances were never observed. These facts support the reliability of the statistical workout [18], also considering that the models were very simple and there were only two latent variables.

Table 4 shows the effects of removing theoretically relevant links from the expanded starting model (eight parameters). The original mismatch values of each model can be found in table 3.

For both fluency types, the link with Semantics was crucial, whereas only for Letter Fluency, the attention load showed a trend, although it fell short of significance. These findings undermine the hypothesis that a poorer category fluency performance is due to the impact of an attentional deficit in AD patients.

Before accepting categorically that attention deficit is not relevant, we must be sure that the conclusion would be the same with a different operational definition of the latent variable Attention. Unfortunately, not all patients of our sample were given additional attention tests; however, for two-thirds of them (35/53) data from the Trail Making Test (TMT) [19] were available. Therefore, we replicated the same LISREL analyses with a different definition of Attention, i.e. no longer defined by the Stroop CWI test *plus* the Digit Cancellation test, but by the Stroop CWI test *plus* the TMT-section B (TMT-B). The latter test requires to connect alternatively, following the natural order of progression, numbers from 1 to 13 and letters from A to N printed on a sheet of paper. TMT-B was preferred to TMT-A because it requires subjects to alternate the search criterion, and we did not include a variable as the difference (TMT-B *minus* TMT-A) because this difference is almost perfectly correlated with TMT-B [20] and is less reliable.

The subsample with a TMT-B evaluation was comparable to the whole sample for age (mean age = 77.2 years, SD = 6.3), education (mean education = 7.1 years, SD = 3.2), and sex

Bizzozero et al.: On Which Abilities Are Category Fluency and Letter Fluency
Grounded? A Confirmatory Factor Analysis of 53 Alzheimer's Dementia Patients

Table 3. LISREL models based on two latent variables (Semantics and Attention) and six observed variables

Latent variables	Observed variables	Basic model	Removing			
			Cat.Fl ↔ Sem	Cat.Fl ↔ Att	Let.Fl ↔ Sem	Let.Fl ↔ Att
Semantics	BORB Association Match test	+0.79	+0.69	+0.77	+0.83	+0.73
	Verbal Semantic Questionnaire	+0.66	+0.77	+0.67	+0.63	+0.68
	Category Fluency task	+0.47	–	+0.53	+0.45	+0.49
	Letter Fluency task	+0.45	+0.43	+0.46	–	+0.60
Attention	Category Fluency task	–0.11 (n.s.)	+0.35*	–	+0.13 (n.s.)*	–0.05 (n.s.)
	Letter Fluency task	–0.24 (n.s.)	+0.25 (n.s.)*	–0.23 (n.s.)	+0.50*	–
	Stroop CWI test (log)	+0.92	–0.85*	+0.90	–0.82*	+0.89
	Digit Cancellation test	+0.61	–0.64*	+0.63	–0.65*	+0.63
Model mismatch (χ^2)		5.120 (d.f. = 6)	11.422 (d.f. = 7)	5.556 (d.f. = 7)	11.699 (d.f. = 7)	7.268 (d.f. = 7)

We report the most comprehensive basic model (with eight parameters) and four reduced models (with seven parameters each) obtained by removing a link between a fluency task and a latent variable in each case (– denotes links removed from the model). Semantics is proportional to semantic competence, but the latent variable Attention, for mere computational reasons, is sometimes proportional to attention impairment, and sometimes (when marked with *) it is proportional to attentional effectiveness – this does not change the meaning of the mismatch statistics (see text, for details). The reported parameters are always different from 0 unless followed by the indication n.s. A higher mismatch corresponds to a worse model. Sem = Semantics; Att = Attention; Cat.Fl = Category Fluency; Let.Fl = Letter Fluency.

Table 4. Relevance of the links originally connecting each latent variable to each fluency type in table 3

Removed link	Mismatch increase (χ^2, d.f. = 1)	p
Between Category Fluency and Semantics	6.302 (11.422 – ref.)	0.012
Between Category Fluency and Attention	0.436 (5.556 – ref.)	n.s. (p = 0.509)
Between Letter Fluency and Semantics	6.579 (11.699 – ref.)	0.010
Between Letter Fluency and Attention	2.148 (7.268 – ref.)	n.s. (p = 0.143)

The mismatch increase refers to the comparison between a 'reduced model' and the most comprehensive original model, corresponding to a χ^2 of 5.120.

(13 males and 21 females). The mean of TMT-B was 363.3 (SD = 113.8, range 122–567). According to the norms [19], pathological subjects were 19/35 (54.3%), a percentage definitely higher than that observed with the Digit Cancellation test. The correlations with the other tests simultaneously included in the second set of LISREL analyses were: –0.208 with Category Fluency, –0.151 with Letter Fluency, –0.221 with the BORB Association Match test, –0.390 with the Verbal Semantic Questionnaire, and 0.421 with the Stroop CWI test.

The second set of LISREL analyses mirrored the first set (table 5). Table 6 illustrates the relevance of the four links that connect the different fluency tests with each latent variable.

The mismatch of the most comprehensive model (χ^2 = 9.490, d.f. = 6, p = 0.148), although still non-significant, was slightly greater with respect to the former analysis (χ^2 = 5.120, d.f. = 6, p = 0.529), probably because the TMT-B sample was smaller. The significance levels were marginal, but showed a strong trend toward the same pattern observed in the former set of

Bizzozero et al.: On Which Abilities Are Category Fluency and Letter Fluency
Grounded? A Confirmatory Factor Analysis of 53 Alzheimer's Dementia Patients

Table 5. Additional LISREL models

Latent variables	Observed variables	Basic model	Removing			
			Cat.Fl ↔ Sem	Cat.Fl ↔ Att	Let.Fl ↔ Sem	Let.Fl ↔ Att
Semantics	BORB Association Match test	+0.67	+0.69	+0.74	+0.73	+0.72
	Verbal Semantic Questionnaire	+0.69	+0.76	+0.69	+0.72	+0.71
	Category Fluency task	+0.44	–	+0.52	+0.28 (n.s.)	+0.44 (p = 0.10)
	Letter Fluency task	+0.44	+0.37 (n.s.)	+0.46	–	+0.59
Attention	Category Fluency task	–0.14 (n.s.)	+0.50	–	+0.26 (n.s.)	–0.10
	Letter Fluency task	–0.22 (n.s.)	+0.25 (n.s.)*	–0.19 (n.s.)	+0.61	–
	Stroop CWI test (log)	+0.79	–0.64*	+0.76	–0.61*	+0.70
	TMT-B	+0.53	–0.64*	+0.55	–0.49*	+0.60
Model mismatch (χ^2)		9.490 (d.f. = 6)	12.018 (d.f. = 7)	9.904 (d.f. = 7)	12.617 (d.f. = 7)	10.240 (d.f. = 7)

See table 3 for the structure of the set of analyses. The reported parameters are always different from 0 unless followed by the indication n.s. (– denotes links removed from the model). In cases marked with *, the latent variable Attention is defined by the computational procedure as attentional effectiveness, in the other cases as attentional deficit – this does not change the meaning of the mismatch statistics. Sem = Semantics; Att = Attention; Cat.Fl = Category Fluency; Let.Fl = Letter Fluency.

Table 6. Relevance of the links originally connecting each latent variable to each fluency type in table 5

Removed link	Mismatch increase (χ^2, d.f. = 1)	p
Between Category Fluency and Semantics	2.528 (12.018 – ref.)	n.s. (p = 0.111)
Between Category Fluency and Attention	0.414 (9.904 – ref.)	n.s. (0.520)
Between Letter Fluency and Semantics	3.127 (12.617 – ref.)	0.077
Between Letter Fluency and Attention	0.721 (10.211 – ref.)	n.s. (0.396)

The mismatch increase refers to the comparison between each 'reduced model' and the most comprehensive original model, corresponding to a χ^2 of 9.490. The significance levels can be considered one-sided.

analyses. The latent variable Attention has definitely little to do with Category Fluency, and even with Letter Fluency. Considering that the sample included in this replication was smaller and here the hypothesis was one-tailed, we feel authorized to conclude that the pattern disclosed in the main set of the LISREL analyses (tables 3, 4) is confirmed.

Summing up, this study shows that both fluency variants mainly depend on the latent variable Semantics; consequently, the very reason why AD patients are more severely impaired on category fluency cannot be explained by the disproportionate impact of attention deficit on category fluency. This negative conclusion, however, does not give a positive answer to the underlying question: if semantics affects both fluency types, why is category fluency more defective than letter fluency in the AD population? Furthermore, why does semantics influence letter fluency to an extent comparable to category fluency, if in the former task the lexical search is guided by word form and not by word meaning?

For a further elucidation of our findings, we considered that the correlation between Category Fluency and Letter Fluency, although moderate (r = 0.322), might have somewhat

Bizzozero et al.: On Which Abilities Are Category Fluency and Letter Fluency
Grounded? A Confirmatory Factor Analysis of 53 Alzheimer's Dementia Patients

Table 7. Models of Category Fluency

	Analysis 1		Analysis 2	
	Semantics	Attention	Semantics	Attention
Category Fluency task	+0.46	–0.14 (n.s.)	+0.43	–0.13 (n.s.)
BORB Association Match test	+0.88		+0.78	
Verbal Semantic Questionnaire	+0.60		+0.67	
Stroop CWI test (log)		+0.91		+0.66
Digit Cancellation test		+0.62		
TMT-B				+0.64

The parameters are different from 0 unless followed by the indication n.s. In all reported analyses, the latent variable Attention is directly proportional to the attention impairment. Analysis 1: attention defined by Stroop CWI test + Digit Cancellation test. Analysis 2: attention defined by Stroop CWI test + TMT-B.

Table 8. Models of Letter Fluency

	Analysis 1		Analysis 2	
	Semantics	Attention	Semantics	Attention
Letter Fluency task	+0.45	–0.23 (p = 0.08)	+0.42	–0.19 (n.s.)
BORB Association Match test	+0.68		+0.65	
Verbal Semantic Questionnaire	+0.79		+0.81	
Stroop CWI test (log)		+0.89		+0.72
Digit Cancellation test		+0.63		
TMT-B				+0.59

In all reported analyses, the latent variable Attention is directly proportional to the attention impairment. For a borderline estimate, the p value is reported. Analysis 1: attention defined by Stroop CWI test + Digit Cancellation test. Analysis 2: attention defined by Stroop CWI test + TMT-B.

blurred the meaning of the links between the single fluency tasks and the latent variables. Hence, we analysed the two fluency variables one at a time against the background of the two semantic variables and of the two attentional variables. The new analyses almost perfectly reproduced the results of the more comprehensive models, and this held true when Attention was defined by both the Stroop CWI test *plus* the Digit Cancellation test, or by the Stroop CWI test *plus* the TMT-B. Tables 7 and 8 show the estimates of this new set of analyses. At this point, we believe that Category Fluency does not depend on Attention; however, it is still not clear why Letter Fluency depends on Semantics.

Tables 7 and 8 offer further useful suggestions for the interpretation of our findings. One is that, although the relevance of Semantics is predominant for both variables, the attentional load of Letter Fluency seems more robust than that of Category Fluency: the parameter corresponding to the former link reaches a 0.08 significance level in the main analysis, where Attention is defined by the Stroop CWI test *plus* the Digit Cancellation test. A second interesting point is that the links between the two semantic variables (the BORB Association Match test and the Verbal Semantic Questionnaire) and the latent variable Semantics were not perfectly parallel between the Category Fluency and Letter Fluency models. For Category Fluency, the link with Semantics tends to be stronger for the BORB Association Match test

Bizzozero et al.: On Which Abilities Are Category Fluency and Letter Fluency Grounded? A Confirmatory Factor Analysis of 53 Alzheimer's Dementia Patients

than for the Verbal Semantic Questionnaire, whereas the opposite disproportion was observed for Letter Fluency. This disproportion should be considered only as a trend, but it might all the same shed light on why Letter Fluency presents a significant link with Semantics, and this possibility will be discussed below.

Discussion

The worse performance of AD patients in Category Fluency as compared to Letter Fluency (56.7 vs. 41.5% of pathological scores) replicates independent data [1], where 65% of the AD patients were pathological on Category Fluency and 49% on Letter Fluency. The main question was to understand why, among AD patients, category fluency is more severely impaired than letter fluency. We had anticipated two alternative explanations: the first based on a disproportionate impairment of semantics over attention, associated to a strong link between category fluency and semantics, the second based on a disproportionate impairment of attention over semantics, with a strong link between category fluency and attention.

The first sub-question, i.e. which domain is more severely impaired in AD patients, cannot be given a definite answer in this study. The two tests defective in more than 50% of the patients were the Verbal Semantic Questionnaire (a semantic measure) and the Stroop CWI test (an attention measure). Within both domains, some measures were probably more sensitive than others; however, we found a good correlation between tests belonging to the same cognitive realm, and we feel it justified to discuss the outcome of the LISREL analyses that are based on structural correlations between different constructs.

As Category Fluency was exempt from an attentional-executive load, our findings are not compatible with the attention deficit hypothesis anticipated in the introduction. It remains to be understood why the impact of Semantics on Letter Fluency was as strong as that acting on Category Fluency. This is rather strange, as letter fluency calls for a search based on word form and not on word meaning. The final set of analyses, reported in tables 7 and 8, may offer a tentative explanation. In the analysis of Category Fluency, the latent variable Semantics was relatively more related to the BORB Association Match test than to the Verbal Semantic Questionnaire; the opposite pattern applies to the Letter Fluency analysis. This might suggest that, while defining the link between Letter Fluency and Semantics, the prevailing role is played by the Verbal Semantic Questionnaire, a verbal task designed to tap semantics, but which certainly requires also a general lexical competence. On the contrary, in the link between Category Fluency and Semantics, the prevailing role is played by the BORB Association Match test, a non-verbal task that might be less sensitive to the level of lexical skills. In other words, the relationship between the latent variable Semantics and Letter Fluency apparently shows a higher load of lexical ability than that found in the relationship between Semantics and Category Fluency. This might account for the relevance of the latent variable Semantics in Letter Fluency.

A final comment concerns the question whether the results of this class of experiments bear generally valid conclusions, or should be viewed as limited to a given population. The mechanisms used by different pathological groups for performing the same task might not be coincident, and the same set of tasks might yield different correlation matrices for different groups. For instance, Basso et al. [21] found that the same semantic associative test (the Colour-Figure Matching test) was significantly more correlated with the Weigl Sorting Test than with the Raven's Progressive Matrices test among non-fluent aphasics, whereas the opposite outcome applied to fluent aphasia patients. In addition, the performance on each fluency test might depend on different basic deficits in the different pathological groups. This point is certainly relevant for the interpretation of the fluency tasks of traumatic brain injury

Bizzozero et al.: On Which Abilities Are Category Fluency and Letter Fluency Grounded? A Confirmatory Factor Analysis of 53 Alzheimer's Dementia Patients

patients who present a disproportionate impairment of letter fluency with respect to semantic fluency [1]. An interesting empirical question to address in future studies is whether the basic components of letter fluency and category fluency impairment (i.e. the relative weight of semantic vs. attentional deficit) are the same in brain injury patients and AD patients.

Conclusions

In conclusion, we can give at least one clear answer to our original questions, indicating that category fluency has not much to do with attention, at least when the role of attention is compared to that of semantics. The reason why AD patients are more impaired on category fluency than on letter fluency remains unknown; however, the analyses carried out separately for Category Fluency and Letter Fluency (tables 7, 8) show a trend toward a heavier role of attention within the models of Letter Fluency. Our findings suggest further exploration in the relationship between attention and letter fluency, but for the time being, and for clinical purposes, the bare conclusion that in AD patients letter fluency taps attention whereas category fluency taps semantics would not seem justified.

Acknowledgements

Disclosure Statement

No conflicts of interest are declared.

References

1 Capitani E, Rosci C, Saetti MC, et al: Mirror asymmetry of Category and Letter fluency in traumatic brain injury and Alzheimer's disease. Neuropsychologia 2009;47:423–429.
2 Robinson G, Shallice T, Bozzali M, et al: The differing roles of the frontal cortex in fluency tasks. Brain 2012; 135:2002–2014.
3 Reverberi C, Laiacona M, Capitani E: Qualitative features of semantic fluency in mesial and lateral frontal patients. Neuropsychologia 2006;44:409–478.
4 Faust ME, Balota DA, Multhaup KS: Phonological blocking during picture naming in dementia of the Alzheimer type. Neuropsychology 2004;18:526–536.
5 Nebes RD, Martin DC, Horn LC: Sparing of semantic memory in Alzheimer's disease. J Abnorm Psychol 1984; 93:321–330.
6 Nebes RD: Semantic memory in Alzheimer's disease. Psychol Bull 1989;106:377–389.
7 Schwartz S, Baldo J, Graves E, et al: Pervasive influence of semantics in letter and category fluency: a multidimensional approach. Brain Lang 2003;87:400–411.
8 Chenery HJ, Murdoch BE, Ingram JCL: An investigation of confrontation naming performance in Alzheimer's dementia as a function of disease severity. Aphasiology 1996;10:423–441.
9 Joereskog KG, Soerbom DD: LISREL 8: User's Reference Guide. Mooresville, Scientific Software Inc, 1993.
10 Brazzelli M, Capitani E, Della Sala S, et al: A neuropsychological instrument adding to the description of patients with suspected cortical dementia: the Milan overall dementia assessment. J Neurol Neurosurg Psychiatry 1994;57:1510–1517.
11 Cazzaniga R, Francescani A, Saetti MC, et al: How to calculate a MMSE-score from a MODA-score (and vice-versa) in Alzheimer's patients. Neurol Sci 2008;24:261–267.
12 Laiacona M, Barbarotto R, Trivelli C, et al: Dissociazioni semantiche intercategoriali: descrizione di una batteria standardizzata e dati normativi. Arch Psicol Neurol Psichiatr 1993;54:209–248.

Bizzozero et al.: On Which Abilities Are Category Fluency and Letter Fluency Grounded? A Confirmatory Factor Analysis of 53 Alzheimer's Dementia Patients

13   Riddoch MJ, Humphreys GW (1993): BORB – The Birmingham Object Recognition Battery. Hove, Lawrence Erlbaum, 1993.

14   Caffarra P, Vezzadini G, Dieci F, et al: Una versione abbreviata del test di Stroop: dati normativi nella popolazione italiana. Nuova Riv Neurol 2002;12:111–115.

15   Della Sala S, Laiacona M, Spinnler H, et al: A cancellation test: its reliability in assessing attentional deficits in Alzheimer's disease. Psychol Med 1992;22:885–901.

16   Capitani E, Laiacona M: Classification and modelling in neuropsychology: from groups to single cases; in Boller F, Grafman J, Rizzolatti G (eds): Handbook of Neuropsychology, ed 2. Amsterdam, Elsevier, 2000.

17   Barbey AK, Colom R, Solomon J, et al: An integrative architecture for general intelligence and executive function revealed by lesion mapping. Brain 2012;135:1154–1164.

18   Hayduk LA: Structural Equation Modelling with LISREL. Essentials and Advances. Baltimore and London, The Johns Hopkins University Press, 1987.

19   Giovagnoli AR, Del Pesce M, Mascheroni S, et al: Trail Making Test: normative values from 287 normal adult controls. Ital J Neurol Sci 1996;17:305–309.

20   Capitani E, Laiacona M, Barbarotto R, et al: How can interference be evaluated in attentional tests? A study based on bi-variate non-parametric tolerance limits. J Clin Exp Neuropsychol 1999;21:216–228.

21   Basso A, Capitani E, Luzzatti C, et al: Different basic components in the performance of Broca's and Wernicke's aphasics on the Colour-Figure Matching Test. Neuropsychologia 1985;23:51–59.