

VRT (verbal reasoning test): a new test for assessment of verbal reasoning. Test realization and Italian normative data from a multicentric study

Benedetta Basagni¹ · Claudio Luzzatti^{2,3} · Eduardo Navarrete⁴ · Marina Caputo⁵ · Gessica Scrocco⁶ · Alessio Damora⁷ · Laura Giunchi¹ · Paola Gemignani⁸ · Annarita Caiazzo⁹ · Maria Grazia Gambini¹⁰ · Renato Avesani¹⁰ · Mauro Mancuso⁷ · Luigi Trojano¹¹ · Antonio De Tanti¹

Received: 22 August 2016 / Accepted: 9 January 2017 / Published online: 17 January 2017
© Springer-Verlag Italia 2017

Abstract Verbal reasoning is a complex, multicomponent function, which involves activation of functional processes and neural circuits distributed in both brain hemispheres. Thus, this ability is often impaired after brain injury. The aim of the present study is to describe the construction of a new verbal reasoning test (VRT) for patients with brain injury and to provide normative values in a sample of healthy Italian participants. Three hundred and eighty healthy Italian subjects (193 women and 187 men) of different ages (range 16–75 years) and educational level (primary school to postgraduate degree) underwent the VRT. VRT is composed of seven subtests, investigating seven different domains. Multiple linear regression

analysis revealed a significant effect of age and education on the participants' performance in terms of both VRT total score and all seven subtest scores. No gender effect was found. A correction grid for raw scores was built from the linear equation derived from the scores. Inferential cut-off scores were estimated using a non-parametric technique, and equivalent scores were computed. We also provided a grid for the correction of results by z scores.

Keywords Verbal reasoning · Assessment · Italian normative data · Brain injury

Introduction

Verbal reasoning is a skill that characterizes and distinguishes human beings, and can be defined as the ability to draw inferences from given information [1]. This complex, multicomponential function implies involvement of various cognitive abilities, such as language, attention, working memory and abstraction, as well as categorization skills. Verbal reasoning ability is acquired gradually during language and abstract thought acquisition, and is completed in early adulthood with maturation of the underlying functional and anatomical substrates (e.g., [2] for development of the capacity to understand metaphoric language).

It has long been known that brain injury of diverse aetiology can be associated with verbal reasoning difficulties [3–5]. Several studies have demonstrated involvement of different areas of both cerebral hemispheres in verbal reasoning tasks. fMRI studies show activation of combinations of brain regions in the frontal, parietal, temporal, and occipital lobes, basal ganglia, and areas of the cerebellum [6]. The parietal cortex seems to have a critical role in resolving transitive inferences [7]. Deficits in deductive

✉ Benedetta Basagni
benedetta.basagni@centrocardinalferrari.it

¹ Centro Cardinal Ferrari, via IV Novembre, 21, Fontanellato, 43012 Parma, Italy

² Dipartimento di Psicologia, Università di Milano-Bicocca, Milan, Italy

³ Milan Center for Neuroscience, Milan, Italy

⁴ Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova, Padova, Italy

⁵ Auxilium Vitae, Volterra, Pisa, Italy

⁶ Università di Padova, Padova, Italy

⁷ Centro di Riabilitazione di Terranuova Bracciolini, Monteverchi, Arezzo, Italy

⁸ Fondazione Don Carlo Gnocchi, Levante Ligure, La Spezia, Italy

⁹ Istituto di Riabilitazione Santo Stefano, Ascoli Piceno, Italy

¹⁰ Ospedale Sacro Cuore Don Calabria, Negrar, Verona, Italy

¹¹ Dipartimento di Psicologia, Seconda Università di Napoli, Caserta, Italy

reasoning have been reported in cases of left focal temporal lobe lesions [8, 9] and left frontal lobe lesions [10–12]. The right frontal gyrus and the right anterior insula seem to be involved in conceptualization tasks [13]. The bilateral prefrontal ventromedial regions, the right orbitofrontal cortex, the medial prefrontal cortex, and the anterior cingulate cortex also seem to be involved [14]. Indeed, frontal lobes contribute to verbal reasoning by integrating and analyzing information and its appropriate use in relation to context [15]. Many studies have also shown high correlation between working memory and verbal reasoning task performance in healthy subjects [12, 16, 17]. Furthermore functional imaging studies suggest that the prefrontal cortex is crucial for analogical reasoning [18].

The first studies to test verbal reasoning were probably by Aleksandr Luria [19] (published in 1976, but the study was originally conducted at the end of the 1920s), who used “odd one out” tasks to investigate classification and abstraction capacities. There are many subsequent studies using single tasks of verbal reasoning to correlate function with neuronal substrates [e.g., 11, 12], but few have attempted to produce standardized material useful for assessing patients with acquired brain injury in clinical practice. Although activation of brain regions is widespread during verbal reasoning tasks, and impaired verbal reasoning is subsequently a frequent and disabling consequence of brain injury, few standardized tests to assess this function are available to clinicians. In batteries for assessing executive functions, such as FAB [20], and in certain IQ tests, such as WAIS-IV [21], some subtests assess the capacity to perceive conceptual relations between words. Although a recent systematic review [22] recognized the WAIS-IV similarity subtest as a valid task for measuring and assessing verbal abstract reasoning, used as a single task, it might be insufficient for assessing this complex function. The recent Family Relation Reasoning Test [23], proposed in the German language, assesses several cognitive operations, such as inference, working memory, and deduction, but it is not available in Italian. The “*Giudizi Verbalì*” test [24] has many limits: its normative data have not been updated and the test only provides norms for individuals aged 40 years and over. Furthermore, young people have difficulty in the Proverb subtest, because proverbs are no longer current in everyday language.

The present study was designed to construct a new test for verbal reasoning, suitable also for young individuals. In the first part of the paper, we describe how we constructed the test, and in the second, we provide normative data for the total score and the seven domain subscores in a sample ($N = 380$) of cognitively healthy participants covering a wide age (16–75 years) and education range. We used consolidated analytical procedures [25] to generate a correction grid for raw scores taking the influence of the main

socio-demographic variables (gender, age, and education) into account, and to transform adjusted scores into equivalent scores (ES) [24, 26]. Moreover, grids for z scores are proposed to facilitate comparison of scores obtained either on the entire test or on the seven subtests.

Methods

Phase 1: development of the verbal reasoning test (VRT) and pilot study

Seven subtests were designed, namely, absurdities, intruders, relationships, differences, idiomatic expressions, family relations, and classifications. The different subtests were identified selecting tasks that have different aspects of verbal reasoning as a common denominator. Ten items plus a warming up item were created for each subtest. The stimuli were constructed with different degrees of difficulty, determined by different aspects, such as degree of abstraction or amount of working memory involved.

Absurdities This test consists of sentences containing conflictual information. The subject has to identify the logical incongruence (e.g., “Outside the farm there was a bright sunshine, while inside it was raining”).

Intruders In this test, the participant has to identify the “intruder” among four words (e.g., “physician, *hospital*, dentist, nurse”).

Relationships This task requires the participant to identify the relationship between a pair of terms and to express the same type of relationship between two other words (e.g., “The relationship between *cold* and *hot* is the same of that between *open* and...”).

Differences In this task, the subject is asked to identify the main characteristic distinguishing two objects or concepts (e.g., “What is the main difference between eye and ear?”).

Idiomatic Expressions This test requires the subject to explain the meaning of certain common idiomatic expressions (e.g., “What does it mean: lift your elbow?”).

Family Relations The participant is asked to specify the degree of familial relationship between relatives in a statement (e.g., “Lucy and Mary are sisters. Mary has a daughter, Anne. What kind of family relation is there between Lucy and Anne?”).

Classifications In this task, the participant has to determine the category to which triplets of words belong (e.g., “What are Milan, Rome and Naples?”).

Participants

107 Italian participants were included in a preliminary sample. They were healthy and with no brain injury,

depression, alcohol and/or drug abuse, severe medical conditions (e.g., neoplasms and organ failure), stroke, or clinically evident cognitive disability. We used a normal score on Digit Span forward [27], as inclusion criterion to ensure normal short-term verbal memory. These healthy Italian volunteers were distributed across age classes (mean age was 45.5 years, SD 16.9 years, range 16–75 years), gender (56 women and 51 men), and education levels (primary school to university, mean formal education was 14 years, SD 5.27 years). Subjects under 45 years of age or with less than 8 years of education were excluded. Informed consent was obtained from all individuals participating in the study.

Procedure

Participants were individually assessed with the forward version of the Digit Span test [27] and the VRT. VRT was administered using the following procedures: (a) items were presented orally and (b) items could be repeated once on request of the participant.

Analyses

Participants' responses on all seven tasks were scored 2, 1, or 0: 2 for a correct response, 1 for a partially correct response, for a concrete example without any elaboration or for a correct answer when the item was repeated after the subject's request, and 0 for either a wrong response or for repetition of the item without any elaboration. Three experienced neuropsychologists independently assessed the responses to each item, comparing scores in the case of disagreement. A scoring manual was drawn up using examples (manual available at: <http://vrt.sstefano.it>). We selected seven of the ten original items for each subtest, excluding items with too high response variability, too low mean response rate (<1), and fewer than 55% of responses rated with a score of 2. A final version of the test was then created, composed of seven subtests, each of them containing seven items (total 49 items). The test can be downloaded from the website: <http://vrt.sstefano.it>.

Phase 2: Collection of normative data for an Italian population

Participants

To obtain normative data, we recruited another 290 healthy participants with the same inclusion criteria and socio-demographic characteristics as for the pilot study. Six centers ranging from Southern to Northern Italy were involved in

the study to ensure representativity across regional Italian variants. Thus, final normative data were collected on a total sample of 397 healthy volunteers distributed across age classes (mean 45.9, SD 17.0, range 16–75 years), gender (204 women and 193 men), and educational levels (from primary school to university, expressed as years of formal education, including post-grad education and/or specialization courses; mean = 13.1 years, SD 4.7, range 3–29). Participants were native Italian speakers without a history of neurological deficits. Informed consent was obtained from all individuals participating in the study.

Procedure

The same procedure as in Phase 1 was used here. Responses were scored by the operators of each center using the correction manual created by the authors at the end of the pilot study.

Analyses

Two different sets of analyses were carried out separately on VRT scores and on scores of each of the seven subtests: (a) equivalent score procedure and (b) Z-score procedure. Statistical analysis was performed with the R program [28].

(a) *Equivalent score procedure* Total VRT score and subtest scores were analyzed by simultaneous multiple regressions to check the influence of the demographic variables age, education, and gender. We first did this by means of linear regression on raw score data. Linear regression was significant for age and education but not for gender, so gender was not considered in the subsequent analyses (see below). Age and education were entered in several multivariate linear regressions to partial out any overlapping effect. We applied the transformations suggested by Capitani and Laiacona [25] to age and education, namely, the logarithmic transformation of age [$\ln(100-\text{age})$] and the square root transformation of education. Four regression models were employed: in the models, age and education were both included as raw values, or as transformed values, or one raw and the other one transformed, alternatively. Of the four models, the model with the best R^2 value was selected. To compare the adequacy of different models, we applied the Akaike information criterion (AIC) [29] and the one with the lowest AIC was selected as the best regression model (see [30] for same approach).

In the following step, from the best fit model, we drew the equations that allow to adjust scores for age and education (separately for VRT total score and for those on the subtests). They were used to standardize all raw values and

to build the tables giving correction values for age and education ranges. Since gender does not have any effect on participants' performance, this variable was not included in the best fit model. The resulting correction grid allows immediate adjustment of raw performance scores of newly tested individuals according to age and education. Correction factors were calculated for four age ranges (16–30, 31–45, 46–60, and 61–75 years) and four educational ranges (in terms of years of formal education: 3–7, 8–12, 13–15, and >15 years). Finally, reference limits were computed by analyzing the whole sample of age- and education-corrected values. The cutoff for each index was computed by solving Wilks's [31] integral equations for 95% tolerance limits at 95% confidence levels. The cut-off value separates pathological performance from normal performance and defines the values corresponding to the equivalent score = zero. According to the equivalent score method, scores were classed into five ranges corresponding to five categories (0, 1, 2, 3, and 4). An equivalent score of 4 indicates above-median performance (>50 percentile ranks), while the equivalent scores of 1, 2, and 3 partitions the intermediate range (between cut-off and median values), according to specific percentile ranks [24].

(b) *Z-score procedure* We calculated the *z* scores for all subtest scores and for the total score. A correction grid was constructed on the basis of the scores.

Results

Seventeen subjects were excluded on the basis of pathological Digit Span scores. The final sample was composed of 380 participants (mean age = 45.9 years, SD 17, range 16–75; mean education = 13.2 years, SD 4.7, range 3–29). The distribution of the sample by gender, age, and education ranges is reported in Table 1. The mean raw score of the whole sample was 82.9 ± 11.7 (range 25–98). The mean raw score of the subtests were the following: Absurdities 11.5 ± 2.5 (range 1–14), Classifications 12.9 ± 1.6 (range 5–14), Differences 12.4 ± 2 (range 3–14), Idiomatic Expressions 11.5 ± 2.1 (range 2–14), Intruders 12.2 ± 2.4 (range 2–14), Relationships 11.8 ± 2.7 (range 2–14), and Family Relations 10.5 ± 2.9 (range 0–14).

(a) *Equivalent score procedure* Linear regression analysis of VRT total score and those for the single subtest showed a significant effect of age (all *ps* < 0.024) and education (all *ps* < 0.001). No significant effects were found for the variable gender (all *ps* > 0.429). The Akaike information criterion (AIC) indicates that the best model is that based on transformed age and education for six scores (Absurdity, Intruder, Relationship, and Difference subtests and the Total scores), on raw age and transformed education for two subtests (Idiomatic Expression and Family Relation subtests).

Table 1 Distribution of the sample by gender, age, and education levels

Age (years)	Education (years)								Total
	3–7		8–12		13–15		>15		
	Females	Males	Females	Males	Females	Males	Females	Males	
16–30	–	–	14	11	15	18	20	16	94
31–45	–	–	13	13	7	15	21	17	86
46–60	7	7	17	12	16	13	15	17	104
61–75	13	12	13	12	10	12	12	12	96
Total	20	19	57	48	48	58	68	62	380

Table 2 Values of linear regression models of the best models (see the main text for details)

	Education			Age		
	β	<i>t</i>	<i>p</i>	β	<i>t</i>	<i>p</i>
VRT total score	0.59	14.56	<0.001	0.15	3.76	<0.001
Absurdities	0.46	9.93	<0.001	0.12	2.72	<0.008
Classifications	0.44	9.96	<0.001	0.20	4.62	<0.001
Differences	0.27	5.41	<0.001	0.10	1.97	<0.049
Idiomatic expressions	0.48	10.21	<0.001	0.24	5.28	<0.001
Intruders	0.40	9.00	<0.001	0.25	5.70	>0.001
Relationships	0.49	10.95	<0.001	0.15	3.56	>0.001
Family relations	0.40	8.40	0.001	–0.11	–2.37	>0.019

The table reports standardized beta values (β), *t* values, and *p* values

Best fit analysis yielded significant models for all sets. Statistical data on the best models for each subtest are reported in Table 2. The raw values of VRT and subtest scores corrected according to the equations of the multiple regression models are shown in Table 3, and the equivalent scores for each subtest are reported in Table 4.

(b) *Z-score procedure* Tables 5 and 6 show the z scores for the whole test and the seven subtests. Assuming two standard deviations below the mean score to indicate pathological performance, for such scores to be obtained by only 2% of the population, z scores below -1.96 can be considered pathological and are indicated in the table with an asterisk. The cutoff is shown for the total and for the single subtest scores. We suggest the use of the correction grid (Table 5) to adjust the row score, before transforming raw scores into z score.

Discussion and conclusion

This paper describes a new test designed to assess verbal reasoning abilities in adults with acquired brain injury. We created seven different tasks investigating different aspects of verbal reasoning. We tested this first set of tasks in a pilot study on a sample of 107 healthy subjects, equally distributed between genders, and for age and education ranges. Based on the results obtained on the pilot study, the number of items included in each task was reduced from ten to seven on the basis of the participants' results, and a manual for scoring the results was created.

In the second phase, we recruited the normative sample of 380 participants. Multiple linear regression analysis on the normative sample failed to reveal any gender effect but showed a significant effect of age and education. We, therefore, replicate the results of Spinnler and Tognoni [24] on the “Giudizi Verbali” test, in which sex effect was even not found (see [32] for a comprehensive review about sex differences in cognitive abilities, and [33] for sex differences in verbal reasoning). We then built a grid for correcting raw point scores on the basis of the patients' socio-demographic characteristics, and introduced two different types of analyses of the results: equivalent scores and z scores, to meet two different requirements. Equivalent scores are mainly indicated for discriminating pathological from normal performances, but are not sensitive to changes in a patient occurring over time, as for brain-damaged patients that were tested before and after treatment, or in the case of a follow-up study in patients suffering of degenerative brain damage. Z scores are more sensitive for comparing follow-up performances, for example, of a same individual at different time intervals.

Table 3 Correction grid of raw scores for the entire VRT and its single subtests (adjusted score = raw score – corrected score)

Age (years)	Education (years)			
	3–7	8–12	13–15	>15
VRT total score				
16–30		+0.18	–3.13	–7.70
31–45		+5.41	–1.39	–8.78
46–60	+12.67	+5.14	–0.20	–7.53
61–75	+16.07	+8.64	+2.10	–6.00
Absurdities				
16–30		+0.01	–0.54	–1.30
31–45		+0.89	–0.24	–1.47
46–60	+2.12	+0.86	–0.03	–1.25
61–75	+2.70	+1.46	+0.37	–0.98
Classifications				
16–30		–0.19	–0.51	–0.99
31–45		+0.51	–0.22	–1.03
46–60	+1.42	+0.58	+0.01	–0.77
61–75	+1.94	+1.14	+0.44	–0.46
Differences				
16–30		–0.07	–0.32	–0.67
31–45		+0.40	–0.14	–0.73
46–60	+1.03	+0.42	+0.00	–0.58
61–75	+1.36	+0.77	+0.25	–0.41
Idiomatic expressions				
16–30		+1.21	+0.53	–0.22
31–45		+1.08	+0.12	–0.86
46–60	+1.42	+0.49	–0.26	–1.33
61–75	+1.24	+0.20	–0.73	–1.74
Intruders				
16–30		–0.49	–0.89	–1.51
31–45		+0.60	–0.38	–1.48
46–60	+1.98	+0.82	+0.05	–1.00
61–75	+2.86	+1.78	+0.83	–0.43
Relationships				
16–30		–0.09	–0.70	–1.56
31–45		+0.99	–0.31	–1.72
46–60	+2.45	+1.00	–0.02	–1.42
61–75	+3.19	+1.77	+0.53	–1.04
Family relations				
16–30		–0.09	–0.61	–1.38
31–45		+0.95	–0.22	–1.49
46–60	+2.23	+0.94	+0.02	–1.24
61–75	+2.75	+1.48	+0.36	–1.01

The test may have different applications. First, it is also suitable for young subjects, since it was calibrated on persons from 16 years of age and over. A second advantage

Table 4 Equivalent scores (ES) for adjusted values of the VRT total score and the single subtest scores (age- and education-corrected scores)

Equivalent score	VRT	Absurdities	Classifications	Differences
0	<65.17	<6.75	<9.82	<7.56
1	65.18–71.76	6.76–8.59	9.83–11.21	7.57–9.68
2	71.77–78.64	8.60–10.32	11.22–12.52	9.69–11.72
3	78.65–84.26	10.33–11.72	12.53–13.19	11.73–12.90
4	>84.27	>11.73	>13.20	>12.91
Equivalent score	Idiomatic expressions	Intruders	Relationships	Family relations
0	<7.71	<7.72	<6.54	<4.30
1	7.72–9.13	7.73–9.45	6.55–8.76	4.31–7.20
2	9.14–10.45	9.46–11.26	8.77–10.82	7.21–9.34
3	10.46–11.82	11.27–12.62	10.83–12.19	9.35–10.99
4	>11.83	>12.63	>12.20	>11

Table 5 Z score for each of the seven subtests

Raw score	Z scores						
	Absurdities	Intruders	Relationships	Differences	Idiomatic expressions	Family relations	Classifications
0	−4.584*	−5.008*	−4.401*	−6.269*	−5.614*	−3.585*	−7.890*
1	−4.184*	−4.598*	−4.029*	−5.761*	−5.126*	−3.244*	−7.280*
2	−3.784*	−4.188*	−3.657*	−5.253*	−4.639*	−2.904*	−6.670*
3	−3.384*	−3.778*	−3.286*	−4.746*	−4.151*	−2.564*	−6.060*
4	−2.984*	−3.368*	−2.914*	−4.238*	−3.663*	−2.224*	−5.451*
5	−2.584*	−2.959*	−2.542*	−3.730*	−3.175*	−1.884	−4.841*
6	−2.184*	−2.549*	−2.171*	−3.223*	−2.687*	−1.544	−4.231*
7	−1.784	−2.139*	−1.799	−2.715*	−2.2*	−1.204	−3.621*
8	−1.384	−1.729	−1.427	−2.208*	−1.712	−0.863	−3.012*
9	−0.984	−1.319	−1.055	−1.700	−1.224	−0.523	−2.402*
10	−0.584	−0.909	−0.684	−1.192	−0.736	−0.183	−1.792
11	−0.184	−0.5	−0.312	−0.685	−0.248	0.156	−1.182
12	0.216	−0.090	0.059	−0.177	0.239	0.496	−0.573
13	0.616	0.319	0.431	0.329	0.726	0.836	0.036
14	1.016	0.729	0.802	0.837	1.214	1.176	0.646

*Pathological values

is that it includes different subtests investigating different aspects of reasoning. Correction of eight different scores (the total VRT score and its seven subtest scores) offers the additional advantage of total or partial use of the tool. For example, in the case of patients with severe brain injury, only certain subtests may be applied, whereas for other patients, the administration of the whole protocol is indicated.

Although we consider that verbal reasoning testing is useful for all patients undergoing neuropsychological evaluation, those with bilateral frontal damage or with left temporal or subcortical lesions of different aetiology

(vascular, traumatic or neurodegenerative) can benefit from more detailed investigation of verbal reasoning capacity.

Further studies are needed to compare the performance of patients in the different subtests. It will be interesting to determine in how much the results correlate between the different subtests and with those of other cognitive tasks and with the site of brain damage. Certain subtests are likely to correlate better with certain cognitive functions (e.g., the Family Relation subtest requires good verbal working memory, whereas the Classification and Intruder subtests are more linked to lexical-semantic and conceptual capacities), and damage to specific brain areas is likely to

Table 6 Z score for the total raw score (RS)

VRT total score													
RS	Z score	RS	Z score	RS	Z score	RS	Z score	RS	Z score	RS	Z score	RS	Z score
0	-7.114*	15	-5.826*	30	-4.539*	45	-3.251*	60	-1.963	75	-0.676	90	0.611
1	-7.028*	16	-5.740*	31	-4.453*	46	-3.165*	61	-1.878	76	-0.590	91	0.696
2	-6.942*	17	-5.654*	32	-4.367*	47	-3.079*	62	-1.792	77	-0.504	92	0.782
3	-6.856*	18	-5.569*	33	-4.281*	48	-2.993*	63	-1.706	78	-0.418	93	0.868
4	-6.770*	19	-5.483*	34	-4.195*	49	-2.908*	64	-1.620	79	-0.333	94	0.954
5	-6.684*	20	-5.397*	35	-4.109*	50	-2.822*	65	-1.534	80	-0.247	95	1.040
6	-6.599*	21	-5.311*	36	-4.024*	51	-2.736*	66	-1.448	81	-0.161	96	1.126
7	-6.513*	22	-5.225*	37	-3.938*	52	-2.650*	67	-1.363	82	-0.075	97	1.212
8	-6.427*	23	-5.139*	38	-3.852*	53	-2.564*	68	-1.277	83	0.010	98	1.297
9	-6.341*	24	-5.054*	39	-3.766*	54	-2.478*	69	-1.191	84	0.096		
10	-6.255*	25	-4.968*	40	-3.680*	55	-2.393*	70	-1.105	85	0.181		
11	-6.169*	26	-4.882*	41	-3.594*	56	-2.307*	71	-1.019	86	0.267		
12	-6.084*	27	-4.796*	42	-3.509*	57	-2.221*	72	-0.933	87	0.353		
13	-5.998*	28	-4.710*	43	-3.423*	58	-2.135*	73	-0.848	88	0.439		
14	-5.912*	29	-4.624*	44	-3.337*	59	-2.049*	74	-0.762	89	0.525		

*Pathological values

affect the results of certain subtests more than others. Overall, this test seems to fill a gap in the range of Italian verbal reasoning tasks.

Acknowledgements We are grateful to Carolina Nozzolillo and Paola Carnevale for their help in the administration of the test.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Johnson-Laird PN (1988) Levels of representation. *Science* 239:1546–1547
- Carriedo N, Corral A, Montoro PR, Herrero L, Ballestrino P, Sebastián I (2016) The development of metaphor comprehension and its relationship with relational verbal reasoning and executive function. *PLoS One* 11(3):e0150289. doi:10.1371/journal.pone.0150289
- Goldstein K (1936) The significance of the frontal lobes for mental performances. *J Neurol Psychopathol* 17:27–40
- Benton AL (1968) Differential behavioral effects in frontal lobe disease. *Neuropsychologia* 6:53–60
- Hirst W, Volpe BT (1988) Memory strategies with brain damage. *Brain Cogn* 8:379–408
- Goel V (2007) Anatomy of deductive reasoning. *Trends Cogn Sci* 11:435–441
- Waechter RL, Goel V, Raymond V, Kruger F, Grafman J (2013) Transitive inference reasoning is impaired by focal lesions in parietal cortex rather than rostralateral prefrontal cortex. *Neuropsychologia* 51:464–471
- Caramazza A, Gordon J, Zurif EB, De Luca D (1976) Right-hemispheric damage and verbal problem solving behavior. *Brain Lang* 3:41–46
- Whitaker HA, Markovits H, Savary F, Grou C, Braun C (1991) Inference deficits after brain-damage. *J Clin Exp Neuropsychol* 13:38
- Adolphs R, Tranel D, Bechara A, Damasio H, Damasio AR (1996) Neuropsychological approaches to reasoning and decision-making. In: Christen Y, Damasio AR, Damasio H (eds) *Neurobiology of decision making*. Springer, New York, pp 157–179
- Goel V, Tierney M, Sheesley L, Bartolo A, Vartanian O, Grafman J (2007) Hemispheric specialization in human prefrontal cortex for resolving certain and uncertain inferences. *Cereb Cortex* 17:2245–2250
- Reverberi C, Shallice T, D'Agostini S, Skrap M, Bonatti LL (2009) Cortical bases of elementary deductive reasoning: inference, memory, and metaduction. *Neuropsychologia* 47:1107–1116
- Kopp B, Rösser N, Tabeling S, Stürenburg HJ, de Haan B et al (2013) Performance on the Frontal Assessment Battery is sensitive to frontal lobe damage in stroke patients. *BMC Neurol* 13:179
- Babcock L, Vallesi A (2015) The interaction of process and domain in prefrontal cortex during inductive reasoning. *Neuropsychologia* 67:91–99
- Petrides M, Pandya DN (1999) Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection. *Eur J Neurosci* 11:1011–1036
- Handley SJ, Capon A, Copp C, Harper C (2002) Conditional reasoning and the Tower of Hanoi: the role of spatial and verbal working memory. *Br J Psychol* 93:501–518
- Capon A, Handley S, Dennis I (2003) Working memory and reasoning: an individual differences perspective. *Think Reason* 9:203–244
- Bunge SA, Wendelken C, Badre D, Wagner AD (2005) Analogical reasoning and prefrontal cortex: evidence for separable retrieval and integration mechanisms. *Cereb Cortex* 15:239–249
- Luria AR (1976) *Cognitive development, its cultural and social foundations*. Harvard University Press, Cambridge

20. Dubois B, Slachevsky A, Litvan I, Pillon BFAB (2000) The FAB A frontal assessment battery at bedside. *Neurology* 55:1621–1626
21. Wechsler D (2008) Wechsler adult intelligence scale-fourth. Pearson, San Antonio
22. Davies G, Piovesana A (2015) Adult verbal abstract reasoning assessment instruments and their clinimetric properties. *Clin Neuropsychol* 29:1010–1033
23. Poinstingl H (2009) The linear logistic test model (LLTM) as the methodological foundation of item generating rules for a new verbal reasoning test. *Psychol Test Assess Model* 51:123–134
24. Spinnler H, Tognoni G (1987) Taratura e standardizzazione italiana di test neuropsicologici. *Ital J Neurol Sci* 6(Suppl 8):1–120
25. Capitani E, Laiacona M (1988) Aging and psychometric diagnosis of intellectual impairment: some considerations on test scores and their use. *Dev Neuropsychol* 4:325–330
26. Capitani E, Laiacona M (1997) Composite neuropsychological batteries and demographic correction: standardization based on equivalent scores, with a review of published data. *J Clin Exp Neuropsychol* 19:795–809
27. Monaco M, Costa A, Caltagirone C, Carlesimo GA (2013) Forward and backward span for verbal and visuo-spatial data: standardization and normative data from an Italian adult population. *Neurol Sci* 34:749–754
28. R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
29. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19:716–723
30. Brugnolo A, De Carli F, Accardo J, Amore M, Bosia LE et al (2016) An updated Italian normative dataset for the Stroop color word test (SCWT). *Neurol Sci* 37:365–372
31. Wilks SS (1941) Determination of sample size for setting tolerance limits. *Ann Math Stat* 12:91–96
32. Halpern DF, Wright T (2000) Sex differences in cognitive abilities, 3d edn. Erlbaum, Hillsdale
33. Colom R, Contreras MJ, Arend I, Garcia Leal O, Santacreu J (2004) *Psychological Record* 54:365–372