




# The Tower of London (ToL) in Italy: standardization of the ToL test in an Italian population

Maddalena Boccia<sup>1,2</sup>  · Dario Marin<sup>3</sup> · Giovanni D'Antuono<sup>1</sup> · Paola Ciurli<sup>2</sup> · Chiara Incoccia<sup>2</sup> · Gabriella Antonucci<sup>1,2</sup> · Cecilia Guariglia<sup>1,2</sup> · Laura Piccardi<sup>4,2</sup>

Received: 21 February 2017 / Accepted: 6 April 2017 / Published online: 21 April 2017  
© Springer-Verlag Italia 2017

**Abstract** Deficit in planning and problem-solving, affecting a wide range of neuropsychological patients, has been widely investigated using the Tower of London (ToL) test, as developed by Shallice (Philos Trans R Soc Lond Ser B Biol Sci 298:199–209, 1). The ToL taps on several executive functions (EF), such as planning, time for planning or rule breaks, which may be usefully indexed by different ToL measurements. However, in its original version, the different aspects involved in ToL are not evaluated in a specific way.

Here, we report the standardization of the ToL, on 896 individuals aged 15–86 years, taking in account individual factors (i.e. gender, age, years of education) which may affect performances on ToL. We computed several indexes on the ToL including score, planning and execution times, perseverations, rule breaks and self-monitoring. We found that these indexes were affected by individual factors such as gender, age and education. Present results not only provide extensive normative data according to gender, as well as different age and education ranges, but also represent a very useful instrument for a more fine-grained diagnosis of EF deficits in a wide

range of neuropsychological patients, including traumatic brain injury and brain-damaged patients, as well as Alzheimer's disease and Parkinson's disease patients.

**Keywords** Tower of London · Executive deficit · Planning · Problem-solving · Frontal lobe · Normative data

## Introduction

The Tower of London (ToL; [1]) is one of the most used tests for assessing executive functions (EF), both in clinical and experimental neuropsychology. In particular, ToL is considered a general measure of visuospatial problem-solving, and more specifically of planning, which results are related to the activity of the prefrontal cortex (see [2, 3] for reviews). The planning skill demands the mental anticipation and evaluation of future actions and of their resulting consequences and it is crucial in everyday life. The main goal of the ToL is to assess the mental planning of a series of moves to match the configuration of beads presented by the examiner. Indeed, in the ToL, individuals are asked to rearrange a given start configuration to reach a final configuration in a predetermined set of moves, also following a set of rules: only one ball can be moved at a time, no ball can be placed outside of the three wooden pegs that have a different length for holding three, two or just one ball, respectively.

In the clinical use, generally, performances are evaluated in terms of accuracy measured as the number of trials correctly solved in the minimal number of moves or response times for the planning and execution phase. Performances result defective in a large range of neurological and psychiatric patients (e.g. Parkinson's disease,

**Electronic supplementary material** The online version of this article (doi:10.1007/s10072-017-2957-y) contains supplementary material, which is available to authorized users.

✉ Maddalena Boccia  
maddalena.boccia@uniroma1.it

<sup>1</sup> Department of Psychology, Sapienza University of Rome, Rome, Italy

<sup>2</sup> Neuropsychology Unit, IRCCS Fondazione Santa Lucia of Rome, Rome, Italy

<sup>3</sup> I.M.F.R. Gervasutta Hospital, Udine, Italy

<sup>4</sup> Department of Life, Health and Environmental Sciences, L'Aquila University, L'Aquila, Italy

stroke, traumatic brain injury, dementia, schizophrenia or attention-deficit/hyperactivity disorder) [4].

Some previous studies focused attention to rule breaks, which occur when subjects fail in complying to the initial rules to solve the ToL's task, since the rule-breaking behaviour seems to be characteristic of some pathologies. Thus, for example, in focal brain-damaged patients, the presence of rule breaks is associated with frontal lesions [5], but rule breaks are also present in neurodegenerative disorders. Both patients with Alzheimer's disease and mild cognitive impairment break rules very often [6–8] and rule breaks increase with the increasing of the problem complexity [6] in patients affected by frontotemporal lobe dementia [9]. Furthermore, the increasing of rule breaks allows to distinguish patients with stable deficits from those with declining cognition [7]. Elevated numbers of rule break also characterize patients with Parkinson's disease [10], schizophrenia [11] and ADHD [12].

In general, performance on ToL correlates with the results of tasks evaluating other cognitive functions, i.e. attention [13], working memory, flexibility, inhibition control [14–17] and fluid intelligence [17–19].

Experimental studies aimed to understand the nature of individual differences and the neural correlates of ToL performances. For example, Boghi et al. [20] underlined the role of gender. In general, they found that brain activity varies with the difficulty of trial, being easier trials associated to neural activity in parietal regions and difficult trials with activity in a wider network including several frontal regions and subcortical structures. Gender differences were also observed suggesting that males and females rely on slightly different cognitive processing for performing ToL. Indeed, males showed higher precuneus activity suggesting relied on visuospatial abilities and females showed higher activity in dorsolateral prefrontal cortex suggesting they relied more on executive processing.

Also, ageing affects performances in ToL [14, 21–23]. Indeed, not only young children but also teenagers may show poorer performances than young adults; furthermore, compared to younger adults, healthy elders have shown decline of performances, which have been interpreted as due to age-induced changes in fluid intelligence [15, 22–24], in visuospatial working memory [14, 15, 24] and also in processing speed [21].

Taken together, these results suggest that not only pathology but also individual factors affect performance on ToL, making mandatory to have normative data allowing to correct raw scores for age, gender and education.

## Materials and method

### Participants

The ToL has been administered to 896 healthy Italian participants aged 15–86 years, who were relatives of inpatients at

the IRCCS Fondazione Santa Lucia or college students at the Sapienza University of Rome. Participants were subdivided into six groups: (1) youngest (Y) included 159 participants aged between 15 and 25 years (education range 5–18 years), (2) young adults (YA) included 146 participants aged between 26 and 35 years (education range 8–18 years), (3) adults (A) included 124 participants aged between 36 and 45 years (education range 5–18 years), (4) middle-aged (MA) included 207 participants aged between 46 and 55 years (education range 5–18 years), (5) older (O) included 168 participants aged between 56 and 65 (education range 5–18 years) and (6) oldest (TO) included 92 participants aged between 66 and 86 years (education range 5–18 years) (for more details about demographics see Table 1).

None of the participants had a history of neurological or psychiatric disease, as confirmed during an informal interview carried out before the test phase. To exclude the presence of deficit on logical and abstract reasoning, all participants performed the Raven's Coloured Progressive Matrices Test [25, 26].

### Tower of London (ToL 16)

The Tower of London test (ToL) includes problems of increasing difficulty. It consists of a board (25 × 10 cm) with three vertical pegs of different increasing length (6, 12, 18 cm) in which are placed three different wooden balls or beads (5 cm diameter) of different colours (red, white and green). The shortest peg only accommodates one bead, the second, two and the third, three.

**Table 1** Demographics, mean, standard deviations and age and education ranges, for male and female participants in each group

Group	Gender	N	Age	Education
Y	F	85	21.08 (2.52)	12.52 (3.30)
Age range 15–25	M	74	19.72 (3.09)	11.49 (3.73)
				Range 5–18
YA	F	71	30.07 (2.89)	14.31 (3.76)
Age range 26–35	M	75	30.81 (2.95)	13.64 (3.68)
				Range 8–18
A	F	71	40.85 (2.88)	11.97 (3.61)
Age range 36–45	M	53	39.92 (2.80)	12.87 (3.69)
				Range 5–18
MA	F	122	50.40 (2.87)	12.56 (3.98)
Age range 46–55	M	85	50.64 (2.68)	13.02 (3.96)
				Range 5–18
O	F	90	60.78 (2.75)	10.84 (3.98)
Age range 56–65	M	78	59.86 (2.52)	11.73 (4.16)
				Range 5–18
TO	F	52	72.21 (5.14)	7.98 (2.98)
Age range 66–86	M	40	71.70 (4.44)	9.85 (3.95)
				Range 5–18

Like the original ToL task [1], the right peg on each trial could only hold one bead, and moving leftward, each successive peg could fit one additional bead.

Subjects are presented with a given bead configuration (starting configuration; Fig. 1a) and a picture of the final configuration (Fig. 1b) they have to reach by moving the three beads, without violating the following rules: (a) solving the problem within a maximum number of moves written on the picture of the final configuration, (b) moving one bead at a time and never holding more than one bead and (c) bear in mind that the smallest peg can contain only one bead, the medium peg can contain two beads and the highest peg can contain all three beads.

Here, we standardized on an Italian sample the version of the test conceived by Shallice [1] following the procedures proposed by Krikorian et al. [27]. Although in these years novel versions of ToL have been proposed (see for example [13]), we decided to adopt this older version, since to our knowledge it is the more diffuse in clinical evaluations. Moreover, we added four new problems to add the original 12 proposed by Shallice in 1982 ([1] see Fig. 1).

The ToL 16 includes 16 trials of increasing difficulty with a maximum number of allowed moves that varies from 2 to 7. For each trial, three attempts are allowed. Following Krikorian

et al. [27], scoring corresponds to the number of solved problems (hereafter called score) according to the number of attempts needed to achieve the solution (i.e. 3 = solved at first attempt; 2 = solved at second attempt; 1 = solved at the third attempt; 0 = not solved).

In solving the trials, participants had to take into account the following rules:

1. Each pegs can accommodate a different number of beads as follows: the first peg (the shortest) one bead, the second (the medium) peg two beads and the third (the longest) three beads.
2. Just one bead might be moved at time. Thus, a bead can be moved only if the previous one has been placed on one of the pegs.
3. The beads cannot be placed outside the pegs (for example, on the table or on the ToL board).
4. The number of allowed moves, which is printed in the bottom part of the picture, must be respected.

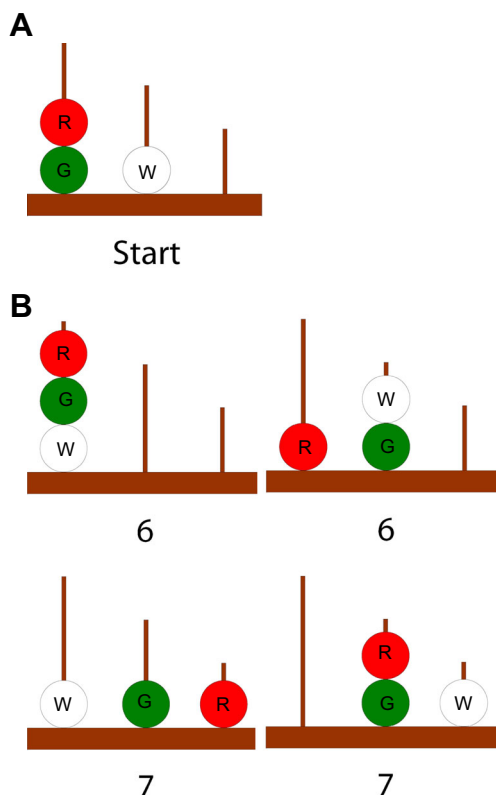
After participants had solved each trial, they were asked whether they believed that their performance was correct or wrong (self-monitoring).

For each item, the experimenter reported on the scoring sheet (see supplementary materials) number of trials necessary for solving the item (maximum 3), the sequence of moves in each attempt, time between instruction and the first move, total time of execution (from the instruction to the end of the item), presence and number of perseveration (repetition of the *first move* of the wrong sequence), presence, number and type of rule breakings.

The following indexes have been calculated: (1) score, as it results from the sum of the scoring on each trial (maximum score = 48); (2) planning time (seconds), as it results from the sum of the time spent on each item between the instruction and the first move; (3) execution time (seconds), as it results from the difference between the sum of the total time spent on each item and the planning time; (4) perseverations, as it results from the sum of the repeated “first move” of the wrong sequence on the same trial; (5) rule breaks, as it results from summing the presence (1 = at least one rule break; 0 = no rule breaks) of rule break on each item and attempt and (6) self-monitoring, as it results from summing the correct evaluation of one’s own performance on each trial (1 = wrong performance recognized as wrong or correct performance recognized as correct; 0 = wrong performance failed recognized as wrong).

## Results

Pearson’s correlation has been calculated to assess whether the performances on the ToL correlated with age, education and gender (point biserial correlation). Age, gender and education were significantly correlated with most of the ToL indexes,



**Fig. 1** a Starting configuration of the Tower of London (ToL) test. b Final configurations and number of allowed moves in the four new trials. G green, R red, W white. Colour figure can be viewed in the online issue

and more specifically with score, perseverations, rule breaks and self-monitoring (Table 2). Furthermore, education significantly correlated with planning time, whereas age and gender significantly correlated with execution time. According to the above reported results, mean and standard deviations (SD) for each index of the ToL are provided in Table 3 according to the age, education and gender. Table 4 reported percentiles of the scores on the ToL.

We also performed a MANOVA to assess the effect of age (Y vs. YA vs. A vs. MA vs. O vs. TO) and gender (male vs. female) on the ToL indexes. Education (a continuous measure) has been included as a covariate. Bonferroni's correction for multiple comparisons has been applied to all post hoc comparisons.

We found a main effect of age on score ( $F_{5,883} = 21.601$ ;  $p < 0.001$ ), planning time ( $F_{5,883} = 2.289$ ;  $p = 0.04$ ), execution time ( $F_{5,883} = 13.265$ ;  $p < 0.001$ ), rule breaks ( $F_{5,883} = 14.926$ ;  $p < 0.001$ ) and self-monitoring ( $F_{5,883} = 20.143$ ;  $p < 0.001$ ). Post hoc comparisons on score showed that TO had worse performances (Y:  $p < 0.001$ ; YA:  $p < 0.001$ ; A:  $p < 0.001$ ; MA:  $p < 0.001$ ; O:  $p = 0.001$ ), followed by O, whose performances were worse than Y ( $p < 0.001$ ), YA ( $p < 0.001$ ), A ( $p < 0.001$ ), MA ( $p = 0.01$ ). MA had worse performances than Y ( $p = 0.02$ ), whereas performances of A, Y and YA did not differ ( $p > 0.05$ ).

Post hoc comparisons on planning time failed in highlighting significant effects, whereas post hoc comparisons on execution time showed that TO had the slowest performances (Y:

$p < 0.001$ ; YA:  $p < 0.001$ ; A:  $p < 0.001$ ; MA:  $p < 0.001$ ; O:  $p < 0.001$ ), followed by O, who were slower than Y ( $p = 0.002$ ). MA were slower than Y ( $p = 0.001$ ), whereas A, Y and YA did not differ ( $p > 0.05$ ).

Post hoc comparisons on rule breaks showed that TO performed more rule breaks than all the other groups (Y:  $p < 0.001$ ; YA:  $p < 0.001$ ; A:  $p < 0.001$ ; MA:  $p < 0.001$ ; O:  $p < 0.001$ ), whereas O performed more rule breaks than A ( $p = 0.003$ ).

Concerning self-monitoring, post hoc comparisons showed that TO performed worse than Y ( $p < 0.001$ ), YA ( $p < 0.001$ ), A ( $p < 0.001$ ) and MA ( $p < 0.001$ ). O performed worse than Y ( $p < 0.001$ ), YA ( $p < 0.001$ ), A ( $p < 0.001$ ) and MA ( $p = 0.002$ ). MA performed worse than Y ( $p = 0.032$ ).

A main effect of gender has been detected for score ( $F_{1,883} = 18.141$ ;  $p < 0.001$ ), execution time ( $F_{1,883} = 5.011$ ;  $p = 0.025$ ), perseverations ( $F_{1,883} = 9.231$ ;  $p = 0.002$ ), rule breaks ( $F_{1,883} = 6.665$ ;  $p = 0.010$ ) and self-monitoring ( $F_{1,883} = 5.418$ ;  $p = 0.020$ ). Females showed lower score and slower execution time than males, made more perseverations and rule breaks and obtained lower self-monitoring scores.

We also found an age by gender interaction on planning ( $F_{5,883} = 2.836$ ;  $p = 0.015$ ) and execution times ( $F_{5,883} = 5.921$ ;  $p < 0.001$ ). Post hoc comparisons showed that males and females differed on planning time in Y ( $p = 0.019$ ) and YA ( $p = 0.020$ ) groups, with females being slower than males in Y but faster than males in YA. The same pattern of results was observed in execution time, with Y females being slower than

**Table 2** Pearson's correlation

		Gender	Age	Education	Score	Planning time	Execution time	Perseveration	Rule breaks	Self-monitoring
Gender	<i>r</i>	1	-.043	.044	.146	.023	-.085	-.100	-.074	.071
	<i>p</i>		.198	.191	.000	.486	.011	.003	.027	.034
Age	<i>r</i>	-.043	1	-.222	-.352	-.020	.229	.115	.266	-.352
	<i>p</i>	.198		.000	.000	.545	.000	.001	.000	.000
Education	<i>r</i>	.044	-.222	1	.284	.137	.055	-.140	-.234	.199
	<i>p</i>	.191	.000		.000	.000	.100	.000	.000	.000
Score	<i>r</i>	.146	-.352	.284	1	.285	.153	-.463	-.409	.624
	<i>p</i>	.000	.000	.000		.000	.000	.000	.000	.000
Planning time	<i>r</i>	.023	-.020	.137	.285	1	.436	-.167	-.187	.115
	<i>p</i>	.486	.545	.000	.000		.000	.000	.000	.001
Execution time	<i>r</i>	-.085	.229	.055	.153	.436	1	-.164	-.116	.099
	<i>p</i>	.011	.000	.100	.000	.000		.000	.000	.003
Perseverations	<i>r</i>	-.100	.115	-.140	-.463	-.167	-.164	1	.487	-.355
	<i>p</i>	.003	.001	.000	.000	.000	.000		.000	.000
Rule breaks	<i>r</i>	-.074	.266	-.234	-.409	-.187	-.116	.487	1	-.482
	<i>p</i>	.027	.000	.000	.000	.000	.000	.000		.000
Self-monitoring	<i>r</i>	.071	-.352	.199	.624	.115	.099	-.355	-.482	1
	<i>p</i>	.034	.000	.000	.000	.001	.003	.000	.000	

**Table 3** Averaged performances (score), planning and execution times (seconds) and standard deviations (SD), for male and female participants in each group. For each group, mean perseverations, rule breaks and self-monitoring have been reported as well

Group	Education	Gender	N	Score	Planning time	Execution time	Perseverations	Rule breaks	Self-monitoring
Y	Low (<8 years) <sup>a</sup>	M	1	32	80	155	3	0	16
		F	24	38.25 (4.67)	147.63 (81.04)	132.08 (44.92)	0.75 (1.36)	1.21 (1.47)	15.29 (1.04)
	Medium (8–12 years)	M	33	41.55 (3.51)	55.45 (88.85)	47.3 (68.59)	0.61 (0.83)	1.85 (1.87)	15.91 (0.29)
		F	47	39.15 (4.37)	206.68 (150.73)	140.87 (31.79)	0.96 (1.22)	1.49 (1.41)	15.49 (0.66)
	High (>17 years)	M	29	41.62 (3.13)	167 (96.28)	127.79 (57.4)	0.66 (0.94)	0.9 (1.05)	15.66 (0.72)
		F	14	40.07 (3.99)	212.64 (118.52)	122.5 (27.35)	1.07 (1)	1.71 (1.73)	15.57 (0.65)
YA	Medium (8–12 years)	M	11	43.82 (3.22)	321.09 (171.22)	160.45 (39.61)	0.27 (0.65)	0.64 (0.81)	15.64 (0.67)
		F	13	37.69 (6.37)	195.85 (122.87)	122.54 (40.88)	0.77 (1.69)	1.54 (1.27)	15.23 (1.09)
	Medium/high (13–17 years)	M	16	40.94 (2.74)	229.56 (114.3)	125.94 (26.94)	0.56 (1.09)	1.88 (1.54)	15.75 (0.45)
		F	27	39.67 (4.45)	182.44 (124.81)	120.26 (27.88)	0.74 (0.98)	1.07 (1.57)	15.56 (0.75)
	High (>17 years)	M	34	40.88 (3.76)	211.38 (154.13)	142.88 (49.48)	0.56 (0.79)	1.12 (1.34)	15.44 (1.08)
		F	31	39.1 (5.41)	194.52 (165.14)	118.1 (27.48)	0.9 (1.04)	1.06 (1.31)	15.35 (1.02)
A	Low (<8 years) <sup>a</sup>	M	25	42.8 (3.58)	274.56 (134.8)	138.79 (44.26)	0.52 (0.96)	1.04 (1.49)	15.88 (0.33)
		F	2	37.5 (0.71)	171.5 (41.72)	137.5 (64.35)	0.5 (0.71)	0 (0)	16 (0)
	Medium (8–12 years)	M	15	40.8 (3.9)	210.13 (148.52)	128.47 (17.89)	0.6 (1.24)	1.2 (1.47)	15.33 (0.9)
		F	23	38 (3.49)	154.65 (78.1)	121.87 (27.19)	0.78 (0.9)	1.7 (2.01)	15.26 (1.39)
	High (>17 years)	M	25	39.88 (3.79)	179.28 (91.2)	127.88 (33.31)	0.68 (0.99)	0.76 (0.83)	15.44 (0.96)
		F	11	39.18 (4.94)	178.73 (69.52)	128.45 (29.15)	1 (1)	0.82 (1.17)	15.45 (1.04)
MA	Low (<8 years) <sup>a</sup>	M	13	40.15 (6.38)	121.62 (53.62)	119.62 (28.7)	1.23 (1.24)	0.69 (0.75)	15.38 (0.77)
		F	1	37	93	145	2	3	14
	Medium (8–12 years)	M	1	42	189	90	0	2	16
		F	43	36.37 (5.77)	162.28 (87.27)	138.66 (54.76)	1.14 (1.66)	2.05 (2.28)	14.93 (1.42)
	High (>17 years)	M	24	37.04 (5.3)	186.04 (124.43)	132.88 (33.19)	1.21 (0.98)	1.92 (1.5)	14.5 (1.64)
		F	45	38.67 (4.36)	160.29 (114.12)	133.67 (33.94)	0.93 (1.32)	1.58 (1.71)	15.2 (1.25)
O	Low (<8 years)	M	34	40.41 (3.02)	210.82 (127.58)	138.18 (39.5)	0.29 (0.58)	1.53 (1.6)	15.56 (0.82)
		F	33	40.45 (4.04)	184.82 (97.3)	150.74 (64.99)	0.52 (0.83)	1.06 (1.32)	15.42 (0.97)
	Medium (8–12 years)	M	26	40.65 (3.68)	164.19 (109.93)	129.46 (35.7)	0.23 (0.51)	1 (1.57)	15.27 (1.25)
		F	7	38.29 (3.86)	253.71 (131.56)	155.43 (54.93)	0.43 (0.79)	0.71 (1.5)	15.71 (0.49)
	High (>17 years)	M	5	36 (2.24)	171.4 (49.96)	141.6 (39.77)	1.2 (1.3)	2.6 (2.3)	15 (1.22)
		F	43	34.37 (6.01)	119.28 (70.15)	131.74 (36.59)	1.3 (1.39)	2.98 (3.08)	13.79 (2.6)
TO	Low (<8 years)	M	29	34.34 (7.54)	150.76 (76.88)	120.1 (34.49)	1.31 (2.04)	2.03 (2.46)	13.9 (2.5)
		F	26	37.96 (5.09)	169.96 (67.59)	144.12 (36.76)	0.62 (0.94)	1.77 (2.3)	14.92 (1.26)
	Medium (8–12 years)	M	27	37.78 (5.94)	160.67 (67.81)	130.87 (31.94)	0.78 (0.75)	2.07 (1.75)	14.78 (1.58)
		F	14	40.21 (3.7)	181 (147.95)	138.5 (26.51)	0.36 (0.63)	1.5 (1.51)	15.14 (1.7)
	High (>17 years) <sup>a</sup>	M	17	39.94 (2.73)	242.12 (182.71)	148.53 (29.97)	0.35 (0.61)	1.24 (1.82)	15.29 (1.16)
		F	17	36.65 (4.06)	202.82 (108.02)	178.35 (71.47)	1.59 (2.09)	3.65 (3.57)	14.82 (1.85)
TO	Low (<8 years)	M	6	39.17 (3.43)	159.33 (43.88)	148.67 (29.86)	0.33 (0.52)	2.5 (1.64)	14.83 (1.47)
		F	26	29.23 (8.5)	170.81 (215.47)	147.04 (60.27)	1.92 (3.07)	5.23 (4.83)	12.5 (3.14)
	Medium (8–12 years)	M	21	32.71 (7)	189 (193.97)	160.43 (72.92)	0.95 (1.16)	3 (3.89)	14.05 (1.63)
		F	8	34.5 (6.97)	161.88 (56.61)	156 (37.26)	1.13 (1.36)	3.13 (2.7)	13.38 (2.92)
	High (>17 years) <sup>a</sup>	M	8	32.25 (8.15)	198 (105.97)	147.88 (47.86)	1 (1.41)	3.88 (3.56)	13.75 (1.98)
		F	1	42	231	141	2	2	16
		M	5	38.2 (3.11)	198.6 (56.36)	209 (43.16)	1 (1.22)	0.8 (0.84)	15.8 (0.45)

<sup>a</sup> Not applicable due to the limited number of participants



**Table 4** Percentiles on the ToL

		Y		YA		A		MA		O		TO	
		F	M	F	M	F	M	F	M	F	M	F	M
Low (<8 years)	5th	<sup>a</sup>	32.00	<sup>a</sup>	<sup>a</sup>	37.00	<sup>a</sup>	37.00	42.00	34.00	33.00	30.00	35.00
	10th		32.00			37.00		37.00	42.00	34.00	33.00	30.80	35.00
	25th		32.00			37.00		37.00	42.00	34.00	34.00	33.00	35.75
	50th		32.00			37.50		37.00	42.00	37.00	36.00	36.00	39.00
	75th		32.00					37.00	42.00	42.00	38.00	40.00	43.00
	90th		32.00					37.00	42.00			42.40	
	95th		32.00					37.00	42.00				
Medium (8–12 years)	5th	27.75	35.40	24.00	36.00	31.40	34.00	25.60	24.50	23.80	19.00	14.00	15.70
	10th	31.50	36.80	26.40	37.40	33.40	35.80	29.40	29.50	27.00	21.00	14.70	22.60
	25th	35.00	39.00	34.00	40.00	35.00	37.00	33.00	34.00	31.00	29.50	23.75	27.00
	50th	39.00	41.00	41.00	40.00	38.00	41.00	38.00	38.00	35.00	36.00	29.00	35.00
	75th	42.00	44.00	42.50	42.00	41.00	45.00	40.00	41.75	39.00	40.00	36.00	38.00
	90th	44.00	46.00	44.80	45.60	42.60	46.00	42.00	42.50	42.20	43.00	40.60	41.00
	95th	44.75	47.30			43.80		43.80	43.75	44.00	44.00	43.95	42.80
Medium/high (13–17 years)	5th	31.40	36.00	30.00	33.75	32.80	31.30	30.30	35.00	26.35	19.00	23.00	20.00
	10th	33.60	38.00	32.40	34.50	33.00	32.60	32.00	36.00	28.40	31.40	23.00	20.00
	25th	36.00	39.50	37.00	38.75	37.00	38.00	36.00	38.00	36.00	37.00	28.25	24.25
	50th	39.00	42.00	40.00	41.00	41.00	40.00	39.00	41.00	40.00	39.00	36.50	32.50
	75th	43.00	44.50	43.00	44.25	42.00	43.00	42.00	43.00	41.25	40.00	41.00	38.00
	90th	45.00	46.00	45.20	46.00	44.00	44.40	44.00	44.50	43.00	43.20		
	95th	45.60	47.00	46.60	46.00	45.40	45.00	44.70	45.00	44.95	44.60		
High (>17 years)	5th	33.00	38.00	28.60	35.00	31.00	25.00	33.00	34.35	32.00	35.00	42.00	35.00
	10th	34.50	38.20	32.20	36.20	31.40	27.40	33.00	35.70	34.00	35.80	42.00	35.00
	25th	36.00	41.00	36.00	40.50	35.00	38.00	38.00	37.75	38.50	38.50	42.00	35.50
	50th	40.50	45.00	39.00	44.00	39.00	42.00	41.00	40.50	40.00	40.00	42.00	38.00
	75th	43.50	46.00	44.00	45.50	44.00	45.00	44.50	43.25	43.25	42.00	42.00	41.00
	90th	45.00	47.60	46.80	46.40	45.60	47.00	45.00	46.00	45.50	43.40	42.00	
	95th			47.40	47.70			45.60	46.65			42.00	

Performances below the lowest percentile (i.e. 5th percentile) may be considered as equivalent score = 0 (i.e. they fall below the 95% of the distribution). Performances above the 50th percentile correspond to equivalent score = 4. Intermediate scores (i.e. those ranging between 5th and 50th percentiles) correspond to 3 intermediates levels (i.e. equivalent scores 1, 2, 3) [28, 29]

<sup>a</sup>Not applicable due to the limited number of participants

Y males ( $p < 0.001$ ) and YA females being faster than YA males ( $p = 0.012$ ).

Education has been found to affect scores ( $F_{1,883} = 36.198$ ;  $p < 0.001$ ) planning time ( $F_{1,883} = 13.279$ ;  $p < 0.001$ ), execution time ( $F_{1,883} = 12.086$ ;  $p = 0.001$ ), perseverations ( $F_{1,883} = 8.272$ ;  $p = 0.004$ ), rule breaks ( $F_{1,883} = 17.037$ ;  $p < 0.001$ ) and self-monitoring ( $F_{1,883} = 9.492$ ;  $p = 0.002$ ) (see Table 2 for the direction of the effects).

## Discussion

Here, we provide the standardization on an Italian sample of a modified version of ToL, obtained by adding four items to

original test by Shallice [1]. It is worth to note that even if several versions of the ToL have been used for clinical practice, there is no previous Italian standardization. Thus, the present study provides the first Italian standardization of the ToL, with several useful indexes, such as the perseveration score, rule break, self-monitoring, the execution and the planning time, which are often reported in experimental studies as characterizing pathological performances.

Overall, group comparisons (MANOVA) support the group sampling we provided here. Actually, we found that participants of different gender and age performed differently. In particular, present results enlighten the presence of significant correlations among several of the ToL indexes and the demographic variables age, gender and education. Thus, as we

reported in the [result](#) section and is evident by a perusal of Table 3, the same score in a certain index should be considered normal or pathological according to the individual age, gender and education. For this reason, we are providing normative data (mean and standard deviation; Table 3), as well as percentiles (Table 4) of the performances subdivided according to the demographic variables.

Concerning the age, we found that all the collected indexes, with the exception of the planning time, correlated with the age. Specifically, score was negatively correlated with age, suggesting that there is a linear relation between increasing age and worse performances on the ToL; indeed, the MANOVA results suggest that score begins to decrease by the age of 46 years (MA had worse scores than Y). Furthermore, as age increased, execution time as well as perseverations and rule breaks increased, while self-monitoring decreases with age increasing. The age-related decreasing in execution time and self-monitoring was observed by the age of 26 years. Instead, rule breaks started to increase by the age of 56 years.

We also found that females differed from males in score and execution time. Noteworthy, Y females (age = 15–25 years) were slower than Y males, but YA (age = 26–35 years) were faster than YA males both in planning time and in execution time, suggesting that by the age of 26, there is an overturning of the gender differences in the speed of planning.

All the collected indexes, with the exception of the execution time, significantly correlated with education, suggesting that it is pivotal to consider such information to compare individual raw scores with education-matched controls.

The influence of demographic variables on ToL performances above discussed deserves some comments. On the one hand, these data are in line with previous observations about the individual differences due to age and gender in planning on the ToL [15, 20–23] and provide new important evidence about the effect of the education. On the other hand, they highlight the need in clinical assessment of comparing performances of neuropsychological patients with those of a group of individuals matched for all of these factors to have a correct classification of patient performance and a more fine-grained picture of his/her EF skills.

As reported in the [introduction](#) section, the ToL has been widely applied in the studies of EF, especially in studies of planning, visuospatial problem-solving, capability to adhere to a set of rules and planning/execution times. Here, we provide the first Italian standardization which takes in account all of these indexes also providing a new index, that is the self-monitoring. Each index provides a picture of a specific EF trait and may be affected by the individual, demographic differences above described. The advantages of present standardization are multiple. First of all, it provides up-to-date normative data about some indexes (i.e. score, execution time) in a large sample of individuals. Secondly, it provides normative

data about indexes such as perseverations and rule breaks which up until now have only been object of clinical. Finally, the presence of multiple indexes may provide a deeper classification and comprehension of the components of planning processing which may be selectively altered in neuropsychological patients, and their quantifications will allow not only to report the presence of a specific disorder (for example, rule breaks) but also to evaluate its variability with time.

In conclusion, present data may be useful to assess EF in a wide range of neuropsychological patients, whose neuropsychological profile has been found to be linked to specific EF deficit. In particular, they may be useful to compare rule breaks of Alzheimer's disease and mild cognitive impairment patients [6–8] or patients with frontotemporal lobe dementia [9]. They may be useful in the assessment of EF in brain-damaged patients, especially those with lesion of the frontal lobe or traumatic brain injury, who have been widely reported to show EF deficits [4]. Self-monitoring index, it may be useful to assess the anosognosia and metacognitive unawareness which affect a wide range of clinical population [30] and are correlated with executive function impairment [31], and it could become a useful index to inform and monitor neuropsychological rehabilitation.

### Limitation

Several groups, such as the youngest groups with less than 8 years of education or the oldest groups with more than 17 years of education, have a limited number of participants, even if this mirrors the actual distribution of education in Italian population. Actually in the last decades due to the educational policy, the lower level of education has been raised from 5 to 8 to 10 years of schooling, and since high school length is 5 years, many individuals go further the compulsory education completing the high school. On the other hand, the low number of participants with more than 17 years of education among the oldest groups may be due to social factors, such as the Second World War (during which many teenagers were forced to interrupt schooling) and the following economic boom of 1950s, including also opening of new industries, which guided personal choices, promoting the job increase instead of higher level of instruction preference. However, the reduced number of individuals in these subgroups makes difficult to draw a definitive picture of the ToL performances for very young individuals with low education as well as for old individuals with high level of education, suggesting some caution in clinical evaluation of their performances on ToL.

### Compliance with ethical standards

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

## References

- Shallice T (1982) Specific impairments of planning. *Philosophical Transactions of the Royal Society of London, Part B* 298:199–209
- Unterrainer JM, Owen AM (2006) Planning and problem solving: from neuropsychology to functional neuroimaging. *J Physiol Paris* 99(4–6):308–317
- Nitschke K, Köstering L, Finkel L, Weiller C, Kaller CP (2017) A meta-analysis on the neural basis of planning: activation likelihood estimation of functional brain imaging results in the Tower of London task. *Hum Brain Mapp* 38(1):396–413. doi:10.1002/hbm.23368
- Sullivan JR, Riccio CA, Castillo CL (2009) Concurrent validity of the tower tasks as measures of executive function in adults: a meta-analysis. *Appl Neuropsychol* 16(1):62–75. doi:10.1080/09084280802644243
- Andrews G, Halford GS, Chappell M, Maujean A, Shum DH (2014) Planning following stroke: a relational complexity approach using the Tower of London. *Front Hum Neurosci* 8:1032. doi:10.3389/fnhum.2014.01032
- Rainville C, Amieva H, Lafont S, Dartigues JF, Orgogozo JM, Fabrigoule C (2002) Executive function deficits in patients with dementia of the Alzheimer's type: a study with a Tower of London task. *Arch Clin Neuropsychol* 17(6):513–530
- Rainville C, Lepage E, Gauthier S, Kergoat MJ, Belleville S (2012) Executive function deficits in persons with mild cognitive impairment: a study with a Tower of London task. *J Clin Exp Neuropsychol* 34(3):306–324. doi:10.1080/13803395.2011.639298
- Sánchez-Benavides G, Gómez-Ansón B, Quintana M, Vives Y, Manero RM, Sainz A, Blesa R, Molinuevo JL, Peña-Casanova J (2010) Problem-solving abilities and frontal lobe cortical thickness in healthy aging and mild cognitive impairment. *J Int Neuropsychol Soc* 16(5):836–845. doi:10.1017/S135561771000069X
- Carlin D, Bonerba J, Phipps M, Alexander G, Shapiro M, Grafman J (2000) Planning impairments in frontal lobe dementia and frontal lobe lesion patients. *Neuropsychologia* 38(5):655–665
- Culbertson WC, Moberg PJ, Duda JE, Stem MB, Weintraub D (2004) Assessing the executive function deficits of patients with Parkinson's disease: utility of the Tower of London-Drexel. *Assessment* 11(1):27–39
- Lam BY, Raine A, Lee TM (2014) The relationship between neurocognition and symptomatology in people with schizophrenia: social cognition as the mediator. *BMC Psychiatry* 14:138. doi:10.1186/1471-244X-14-138
- Riccio CA, Wolfe ME, Romine C, Davis B, Sullivan JR (2004) The Tower of London and neuropsychological assessment of ADHD in adults. *Arch Clin Neuropsychol* 19(5):661–671
- Culbertson WC, Zillmer EA (2005) Tower of London—Drexel University, second edition (TOLDX). Multi-Health Systems, Toronto
- Phillips LH, Wynn V, Gilhooly KJ, Della Sala S, Logie RH (1999) The role of memory in the Tower of London task. *Memory* 7(2):209–231
- Phillips LH, Smith L, Gilhooly KJ (2002) The effects of adult aging and induced positive and negative mood on planning. *Emotion* 2(3):263–272
- Carder H, Handley S, Perfect T (2004) Deconstructing the Tower of London: alternative moves and conflict resolution as predictors of task performance. *Q J Exp Psychol A* 57(8):1459–1483
- D'Antuono G, La Torre FR, Marin D, Antonucci G, Piccardi L, Guariglia C (2016) Role of working memory, inhibition, and fluid intelligence in the performance of the Tower of London task. *Appl Neuropsychol Adult* 20:1–11
- Unterrainer JM, Rahm B, Kaller CP, Leonhart R, Quiske K, Hoppe-Seyler K, Meier C, Müller C, Halsband U (2004) Planning abilities and the Tower of London: is this task measuring a discrete cognitive function? *J Clin Exp Neuropsychol* 26(6):846–856
- Zook NA, Davalos DB, Delosh EL, Davis HP (2004) Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. *Brain Cogn* 56(3):286–292
- Boghi A, Rasetti R, Avidano F, Manzone C, Orsi L, D'Agata F, Caroppo P, Bergui M, Rocca P, Pulvirenti L, Bradac GB, Bogetto F, Mutani R, Mortara P (2006) The effect of gender on planning: an fMRI study using the Tower of London task. *NeuroImage* 33(3):999–1010
- Andrés P, Van der Linden M (2000) Age-related differences in supervisory attentional system functions. *J Gerontol B Psychol Sci Soc Sci* 55(6):373–380
- Bugg JM, Zook NA, DeLosh EL, Davalos DB, Davis HP (2006) Age differences in fluid intelligence: contributions of general slowing and frontal decline. *Brain Cogn* 62(1):9–16
- Zook N, Welsh MC, Ewing V (2006) Performance of healthy, older adults on the Tower of London Revised: associations with verbal and nonverbal abilities. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* 13(1):1–19
- Rabbitt P, Lowe C (2000) Patterns of cognitive ageing. *Psychol Res* 63(3–4):308–316
- Basso A, Capitani E, Laiacona M (1987) Raven's coloured progressive matrices: normative values on 305 adult normal controls. *Funct Neurol* 2:189–194
- Raven JC (1938) Standard progressive matrices: sets A, B, C, D and EHK. Lewis, London
- Krikorian R, Bartok J, Gay N (1994) Tower of London procedure: a standard method and developmental data. *J Clin Exp Neuropsychol* 16(6):840–850
- Spinnler H, Tognoni G (1987) Standardizzazione e taratura italiana di test neuropsicologici. *Ital J Neurol Sci* 8(Suppl):1–120
- Bianchi A, Dai Prà M (2009) Twenty years after Spinnler and Tognoni: new instruments in the Italian neuropsychologist's toolbox. *Neurol Sci* 29:209–217
- Paolucci S, Di Vita A, Massicci R, Trallesi M, Bureca I, Matano A, Iosa M, Guariglia C (2012) Impact of participation on rehabilitation results: a multivariate study. *Eur J Phys Rehabil Med* 48(3):455–466
- Ciurli P, Bivona U, Barba C, Onder G, Silvestro D, Azicnuda E, Rigon J, Formisano R (2010) Metacognitive unawareness correlates with executive function impairment after severe traumatic brain injury. *J Int Neuropsychol Soc* 16:1–9