



RESEARCH ARTICLE

# Normative data for Lezak’s Tinkertoy test in healthy Italian adults [version 1; referees: 2 approved with reservations]

Franca Crippa<sup>1</sup>, Luca Cesana<sup>2</sup>, Roberta Daini<sup>1</sup>

<sup>1</sup>Department of Psychology, University of Milano-Bicocca, Milan, Italy

<sup>2</sup>Professional University School of the Italian Switzerland, Lugano, Switzerland

**v1** First published: 22 Apr 2016, 5:727 (doi: [10.12688/f1000research.8409.1](https://doi.org/10.12688/f1000research.8409.1))  
 Latest published: 22 Apr 2016, 5:727 (doi: [10.12688/f1000research.8409.1](https://doi.org/10.12688/f1000research.8409.1))

**Abstract**

The Tinkertoy test is a tool for the neuropsychological assessment of executive functions and a predictor of employability. Originally a children’s toy comprising pieces to assemble freely, the TinkerToy Test examines organizational abilities, planning, and response flexibility. It allows subjects to use their own initiative and does not force them to choose from a series of predetermined alternatives. Tinkertoy test normative values were collected from 256 neurologically healthy Italian subjects.

Multivariable analysis showed sex and education to have significant confounding effects. Adjusted and inferential cut-off points were determined and converted into equivalent scores, applying a distribution-free technique.

**Open Peer Review**

Referee Status: ? ?

	Invited Referees	
	1	2
<b>version 1</b> published 22 Apr 2016	<b>?</b> report	<b>?</b> report
<b>1 Luigi Trojano</b> , Second University of Naples Italy		
<b>2 Andrea Peru</b> , University of Florence Italy		

**Discuss this article**

Comments (0)

**Corresponding author:** Franca Crippa ([franca.crippa@unimib.it](mailto:franca.crippa@unimib.it))

**How to cite this article:** Crippa F, Cesana L and Daini R. Normative data for Lezak’s Tinkertoy test in healthy Italian adults [version 1; referees: 2 approved with reservations] *F1000Research* 2016, 5:727 (doi: [10.12688/f1000research.8409.1](https://doi.org/10.12688/f1000research.8409.1))

**Copyright:** © 2016 Crippa F *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](https://creativecommons.org/licenses/by/4.0/) (CC0 1.0 Public domain dedication).

**Grant information:** The author(s) declared that no grants were involved in supporting this work.

**Competing interests:** No competing interests were disclosed.

**First published:** 22 Apr 2016, 5:727 (doi: [10.12688/f1000research.8409.1](https://doi.org/10.12688/f1000research.8409.1))

## Introduction

Executive functions are the cognitive capacities that control lower-level functions and are essential to future-oriented thought and behaviour. They are affected by head injuries<sup>24</sup> or arise because of a focal frontal lesion, either cortical<sup>5,24,25</sup> or subcortical<sup>13</sup>. In particular, the term executive functions refers to cognitive, emotional and behavioural aspects of conduct involved in achieving a specific purpose. Executive functions include processes that are complex, mixed together and in constant interaction. They facilitate the optimum adaptation of the individual to the environment<sup>1,16,23,24</sup>. Lezak<sup>16</sup> suggested the division of competencies into four specific components: volition, planning, goal-oriented behaviour and effective performance. Several psychometric instruments are available for evaluating executive functions during neuropsychological examinations. However, most of them are generally highly structured (the task and the stimuli set the goal and the processes required to achieve that goal)<sup>15,16</sup>. Moreover, none of the tools currently in use for evaluating the performance in the domain of executive functions is able to assess how the patients are able to formulate a goal and to plan how to pursue it, which are prerequisites for a return to work as well as for social life.

The Tinkertoy was created originally as a toy for kids made of various wooden and plastic pieces (wooden dowels, knobs, wheels, connectors, caps, points), to be assembled freely in order to make constructions. Based on this toy, Lezak created a test for the neuropsychological assessment of executive functions, which gives subjects the opportunity to use their own initiative and does not force them to choose from a series of predetermined alternatives. In fact, one of the most relevant characteristics of frontal lobe syndrome is an environment-dependent behaviour, which makes it difficult to cope with the requirements of everyday life. In this respect, Lezak's Tinkertoy test (TTT) stands out, because it was specifically conceived to examine the ability to generate the most achievable goal, to organize, to plan and act, and to respond in a flexible way in a given context<sup>15</sup>. At the outset, studies of the TTT showed that it could be considered a useful predictor of employability<sup>3,15</sup>. Particularly, some researchers found the TTT complexity score correlated more positively with the employability of traumatic brain-damaged patients, than other tests for executive functions, such as trail making test -B, maze tracking, and several WAIS-R subtests<sup>9,10</sup>. Ownsworth and Shum<sup>20</sup> showed that the difference in TTT scores between employed and unemployed patients after strokes was highly significant ( $p < 0.005$  in a group of 27 subjects.) According to the authors, the TTT seems to describe productivity outcomes better than other tests of executive functioning (*i.e.* the FAS test and the five -point test), independent of the presence of hemiplegia and the elapsed time since the stroke. Furthermore, the TTT has been shown to be useful both for differentiating between types of dementia and for evaluating the severity of dementia<sup>12,17</sup>. In addition, subjects generally find the TTT interesting or amusing, so that it is easy to carry out this test, even in the cases of patients who are not very cooperative. Despite the fact that the TTT is commonly used in a clinical context, the only normative data refers to a very small sample of non-Italians<sup>16</sup>. Given both the potential relevance of this instrument for neuropsychological practice, and the lack of any validation so far, the present study aimed at setting TTT normative values in Italian adults, in order to determine firstly inferential

cut-off points and their tolerance limits, and then equivalent scores, applying a statistical technique developed for neuropsychological tests<sup>7,8,22</sup>.

## Materials and methods

### Participants

Two hundred and fifty six neurologically unimpaired Italian subjects (mean age: 44.6; sd 20.85, range: 15–86 years) enrolled in this study on a voluntary basis, with verbal consent. The Research Ethics Committee of the University of Milano-Bicocca approved this (permit number: RM-2016-40) as a minimal risk study, whereby a signed consent document was not required. The subjects were nearly equally distributed according to sex (126 women and 130 men) and age class (range: 15 to 86 years). The level of education (from primary school to university) was recorded in years. Nobody showed a history or evidence of psychiatric disorders or dementia. The demographic distribution of the sample is shown in [Table 1](#).

### Stimuli and procedure

Test items, administration procedure and scoring criteria in this study followed the ones described by Lezak in the fourth edition of the Neuropsychological Assessment Handbook<sup>16</sup>. The test items were selected from the classic version of the Tinkertoy set. Namely, 50 items were used: 24 wooden dowels (4 red, 4 green, 4 orange, 6 blue, 6 yellow), 10 wooden knobs, 4 wooden wheels, 4 wooden caps, 4 wooden connectors, and 4 plastic yellow connectors ([Figure 1](#)).

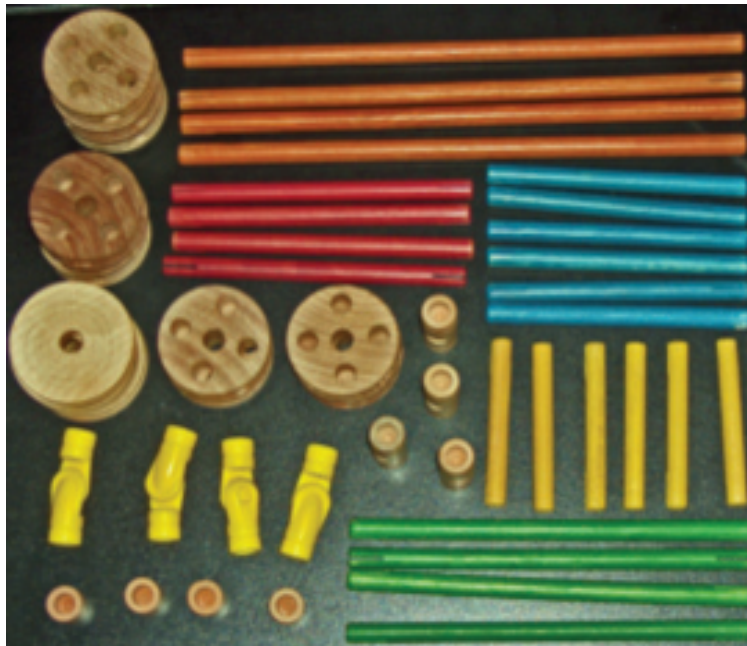
Each subject was individually presented with the aforementioned 50 pieces in different colours and forms, placed at random on a clean surface, and were told to build up whatever construction they wanted with a 5-minute minimum time limit, but no maximum time limit. On completion, the subjects were asked to say what the construction represented. Assessment took into account 7 performance variables: 1. Made construction(s) - whether the subject made any combination of pieces; 2. Number of pieces - total number of pieces used; 3. Name - whether and when the subject gave a name appropriate to the construction's appearance; 4. Mobility (wheels working) and moving parts; 5. Three-dimensionality - whether the subject's construction had three dimensions; 6. Free-standing - whether the subject's construction stayed standing; 7. Errors - pieces forced together (misfit), connections not properly made (incomplete fit), and dropped and not picked up pieces (see [Table 2](#)). At the end a complexity score was given, determined by the sum of the points earned in each of the performance variables, with a maximum of 12 points (for two examples, see [Figure 2](#)).

### Data analysis

The choice of the equivalent scores procedure was prompted by the need to obtain norms that could be directly compared to the already available norms of a wide set of other neuropsychological tests. In the first place, the influence of age, education and gender, the latter dichotomised, was evaluated through a linear multiple regression model, with least square estimation method. Several monotonic transformations of independent variables were analysed and the most effective in reducing the residual variance was adopted. The effect of each variable was studied partialling out the effect held in common with the other variables, after discarding

**Table 1. Distribution of the experimental sample (n=256) according to age and education level.** Values represent the number of subjects.

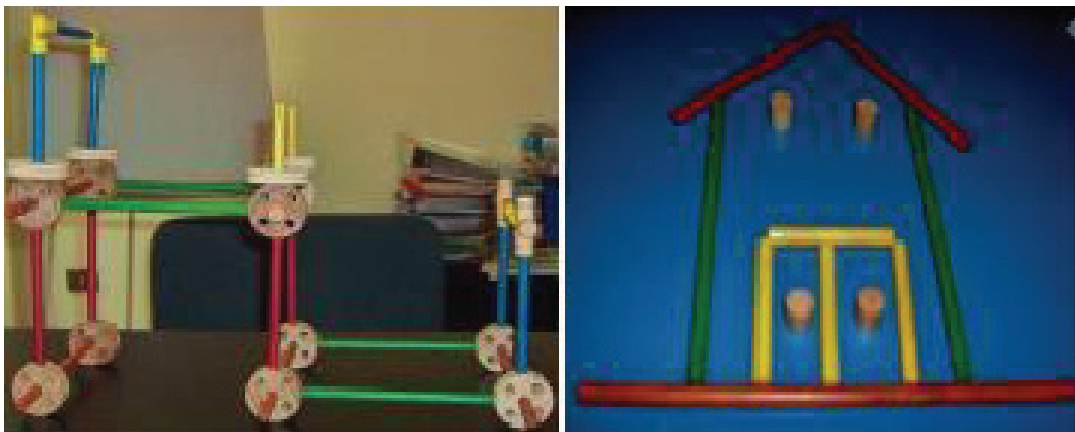
	Age, years						Total
	≤18	19–29	30–44	45–59	60–69	≥70	
<b>Educational Level</b>							
≤5 years							
male	0	0	0	5	6	9	20
female	0	0	0	3	4	20	27
6–8 years							
male	4	8	3	5	6	1	27
female	2	3	5	7	5	1	23
9–13 years							
male	12	6	9	5	5	1	38
female	11	7	4	7	6	3	38
14–16 years							
male	0	6	8	1	1	0	16
female	0	10	2	3	1	1	17
≥ 17 years							
male	0	10	7	6	3	3	29
female	0	8	5	3	3	2	21
Total							
male	16	30	27	22	21	14	130
female	13	28	16	23	19	27	126



**Figure 1.** Tinkertoy items used by Lezak<sup>16</sup> for the Tinkertoy test.

**Table 2. TTT scoring criteria.**

Variables	Scoring Criteria	Points
1. <b>Made construction(s) MD</b>	whether S makes any combination of pieces	1
2. <b>Number of pieces NP</b>	total number of pieces used	<ul style="list-style-type: none"> <li>• 1 = NP ≤20</li> <li>• 2 = NP ≤30</li> <li>• 3 = NP ≤40</li> <li>• 4 = NP ≤50</li> </ul>
3. <b>Name N</b>	whether and when S gives a name appropriate to the construction's appearance	<ul style="list-style-type: none"> <li>• 0 = none N</li> <li>• 1 = description or post hoc naming</li> <li>• 2 = vague or inappropriate N</li> <li>• 3 = appropriate N</li> </ul>
4. <b>Mobility M a) and moving parts M b)</b>	<ul style="list-style-type: none"> <li>• M a): working wheels</li> <li>• M b): moving parts</li> </ul>	<ul style="list-style-type: none"> <li>• M a) = 1</li> <li>• M b) = 1</li> </ul> (maximum: 2 points)
5. <b>Tridimensionality 3D</b>	whether the construction has three dimensions	1
6. <b>Free-standing S</b>	whether the construction stays standing	1
7. <b>Errors E</b>	<ul style="list-style-type: none"> <li>• misfit: pieces forced together</li> <li>• incomplete fit: connections not properly made</li> <li>• dropped and not picked up pieces</li> </ul>	-1 for each error



**Figure 2.** Two examples of performance at the TTT by a neurologically unimpaired subject and a TBI patient recruited exclusively for this comparison. The first is a male, 46 years old, with 13 years of education; his performance has been evaluated as 11.64 (corrected score) and 4 as equivalent score, according to Table 2. The TBI patient is a female, 36 years old, with 13 years of education; her performance has been evaluated as 4.33 (corrected score) and 0 as equivalent score.

age, as non significant as a covariate. In this way, it was possible to estimate the effects of confounding factors on the raw scores and, based on these estimates, adjusted scores were computed, adding or subtracting the contribution of the significant confounding effect. After ranking adjusted scores, Wilks' nonparametric procedure was applied to set tolerance limits<sup>26,27</sup> for pathological TTT result 0

(the lower 5% of all population). The maximum equivalent score, 4, was set with the analogous procedure for the upper 5% of population, whereas equivalent scores 1,2,3 were determined based on the ranking. Spss 21 package for the Social Science led to linear model estimation and to the ranking, Wilks' tolerance intervals by mean of the R package 'tolerance'<sup>27</sup>.

**Results**

**Dataset 1. Raw data for ‘Normative data for Lezak’s Tinkertoy test in healthy Italian adults’, Crippa *et al.* 2016**  
<http://dx.doi.org/10.5256/f1000research.8409.d118684>

Normality criteria are generally appraised by comparing one subject’s performance to that of all the other subjects. This implies homogeneity across the subjects in the comparison, and hence imposes the requirement that all possible factors influencing performance have been taken into account and removed from the raw scores. From a statistical point of view, this aim can be accomplished using stratification, which nonetheless, in some cases, raises problems concerning the sample size in each stratum. Alternatively, the effect of confounding factors can be removed from raw scores in a multiple regression model<sup>4,8,22</sup>. In order to set correction grids for the raw values of participants’ complexity scores, a linear model for the simultaneous effect of sex, age and educational level in years was fitted. Apart from sex, coded as a dummy variable, all dependent and independent variables were centred, where centring each variable on its mean corrected for any overlap with the effect of other terms of the model. The multiple regression proved significant ( $F_{2,242} = 9.08$ ;  $p < 0.001$ , adjusted  $R_{square} = 0.45$ ). With regard to regression coefficients, sex and education proved significant ( $p = 0.002$  and  $p = 0.008$  respectively), whereas age did not, due to multicollinearity with education ( $r_p = 0.422$ ;  $p < .001$ ) and it was therefore discarded. On average, females obtained lower scores than males (8.71 versus 9.40, sd 1.743 and 1.645 respectively). Education played a positive but modest role, an increase in the score from one education class to the adjacent one accounting approximately half a point (Table 3).

Let  $y_f, y_m$  indicate the score of a female and a male respectively and  $x$  the number of years of education. Then, the estimated impact of confounding variables on the TTT Complexity Centred raw scores can be expressed as a linear function of the confounding variables, sex as a dichotomous variable (males coded as 0, females as 1) and centred years of education.

$$(y - \bar{y}) = \beta_0 x_1 + \beta_1 (x_1 - \bar{x}_1) \tag{1}$$

The estimation of the linear regression for the raw scores gives:

$$\begin{aligned} (y_m - 9.06) &= -.06 \times (x - 11.35) \\ (y_f - 9.06) &= .645 - .06 \times (x - 11.35) \end{aligned} \tag{2}$$

Accordingly, adjustment was performed subtracting the estimated contribution of the confounding variables from each raw score, distinctly for females, with  $x = 1$ , and for males, with  $x = 0$  in (2). In order to produce the adjustment to be applied to patients raw scores evaluated in rehabilitation practice, Table 4 shows the correction grid with the points to be added to raw Complexity Scores in order to calculate adjusted scores. Once the adjusted distribution had been computed, the identification of a cut-off point that assessed normality or impairment was a crucial step<sup>19</sup>. The appropriate

**Table 3. Increase in score by education.**

Education in years	Mean score	Standard deviation
≤5	8.64	1.580
6–8	8.78	1.607
9–13	9.22	1.771
14–16	9.03	1.630
≥17	9.50	1.877

**Table 4. Correction grid for raw TTT complexity scores.**

	Education, years				
	≤5	6–8	9–13	14–16	≥17
Sex					
F	+1.06	+0.85	+0.62	+0.38	-0.28
M	-0.25	-0.44	-0.68	-0.88	-1.04

criteria were represented by the interval underlying the lowest 5% tail of the adjusted scores in the cumulative distribution.

However, misclassification of performance may arise and needs to be taken into account. In using the widely accepted value of the lower 5% of the normal population (regarded as a reasonable criterion for classifying subjects that are probably not normal) there is an inherent risk of incorrect categorization. The estimation of inferential tolerance limits enable one to obtain the thresholds above (or below) which there is at least (or at most) a desired percentage of the population, and the estimation of these limits keeps errors in performance assessment under control<sup>7,18</sup>. With the thirteenth observation, corresponding to the value of 6.25, representing the fifth percentile of the cumulative distribution function, the tenth and the sixteenth observation were identified as the outer and the inner limits, yielding the values of 5.86 and 6.44 respectively. Values equal to, or lower than, the outer tolerance limit indicate a pathological performance, with a controlled error risk. In order to compare the performance in this test to those in other tests, the standardization issue needs to be faced. The commonly used z-scores raise various difficulties, such as an alteration of the statistical dispersion of adjusted scores and problems with floor and ceiling values<sup>7</sup>. Distribution-free techniques are required here, since the best standpoint seems to be that of regarding adjusted scores as raw estimates of performance and hence converting them into an ordinal scale with just a few ordinal values, by means of the cumulative function of adjusted scores. A 5-point scale from (0 to 4), termed equivalent scores, is widely used, where 0 indicates the score that lies below the outer non-parametric tolerance limit of adjusted scores, Equivalent scores 1, 2 and 3 are intermediate between 0 and 4, *id est* they are obtained in the cumulative adjusted scores distribution. The equivalent score 4 indicates a performance equal to or superior to the median, thus no longer distinguishing between scores found in the upper half of the distribution.



Equivalent scores 1, 2 and 3 are intermediate between 0 and 4 on a quasi-interval scale. An equivalent score equal to 0 is considered below the normal range, with a controlled error risk. This contracted scale of equivalent scores is then measured on a quasi-interval scale<sup>8</sup> and may be viewed as a standardisation of adjusted scores. [Table 5](#) shows the equivalent score limits, the density (*i.e.* the number of subjects within each equivalent score), and the cumulative frequency of subjects from 0 to 4 equivalent scores.

**Table 5. Equivalent scores.**

Equivalent score	Score interval	Density	Cumulative frequency
0	0 – 5.85	10	10
1	5.86 – 7.02	20	30
2	7.03 – 8.08	41	71
3	8.09 – 9.02	56	128
4	9.03 +	128	256

## Discussion

The TTT has proved to be a highly sensitive instrument for the assessment of organization, planning abilities and response flexibility in a less structured context, compared with other neuropsychological tests generally used to evaluate executive functions<sup>16</sup>. Previously, the TTT could not be administered during psychometric neuropsychological examinations, owing to the absence of normative values, apt to compare the performance of frontal brain-damaged patients with the mean performance of unimpaired subjects with similar demographic features. The present study has filled this gap, establishing TTT normative data for a wide, healthy population sample, representative of the Italian population (n=256). Statistical analyses were adopted according to the methodology that is most widely used in Italy for the computation of normative values<sup>2,6–8,14,21,22</sup>. Our findings showed how TTT performance was affected by sex and education. In particular, males performed better than females and the higher the education the higher the TTT scores. Both effects have already been found in other tests for

executive functions, such as the Wisconsin card sorting and Weigl tests<sup>14</sup> and show an effect of culture and learning in structuring high-level functions. The relationship with education was also found by Apollonio and collaborators with the FAB<sup>3</sup>. Adjusted scores and inferential cut-off scores were calculated. Moreover, adjusted scores were transformed into equivalent scores, since the availability of equivalent scores makes it possible to evaluate whether a patient presents a homogeneous cognitive profile, or if he/she presents selective deficits in one or more cognitive areas. Therefore, it is now possible to compare the performance of brain-damaged patients directly with the TTT and other neuropsychological tests, using normative data with equivalent scores.

## Data availability

*F1000Research*: Dataset 1. Raw data for 'Normative data for Lezak's Tinkertoy test in healthy Italian adults', Crippa *et al.* 2016., [10.5256/f1000research.8409.d118684](https://doi.org/10.5256/f1000research.8409.d118684)<sup>28</sup>

## Consent

Written informed consent was obtained from each patient in the normative study and from a pathological patient whose construction is depicted.

## Author contributions

LC provided expertise in the clinical practice and the administration of the test. RD designed the study and supervised data collection. FC performed the statistical analysis. FC and RD prepared the first draft of the manuscript, LC contributed to the preparation of the manuscript. All authors were involved in the revision of the draft manuscript and have agreed to the final content.

## Competing interests

No competing interests were disclosed.

## Grant information

The author(s) declared that no grants were involved in supporting this work.

## References

- Alvarez JA, Emory E: **Executive function and the frontal lobes: a meta-analytic review.** *Neuropsychol Rev.* 2006; **16**(1): 17–42. [PubMed Abstract](#) | [Publisher Full Text](#)
- Apollonio I, Leone M, Isella V, *et al.*: **The Frontal Assessment Battery (FAB): normative values in an Italian population sample.** *Neurol Sci.* 2005; **26**(2): 108–116. [PubMed Abstract](#) | [Publisher Full Text](#)
- Bayless JD, Varney NR, Roberts RJ: **Tinkertoy test performance and vocational outcome in patients with closed head injuries.** *J Clin Exp Neuropsychol.* 1989; **11**(6): 913–917. [PubMed Abstract](#) | [Publisher Full Text](#)
- Benton AL, Hamsher KdeH, Varney N, *et al.*: **Contributions to Neuropsychological Assessment.** New York, Oxford University Press. 1983.
- Bianchi L: **The functions of the frontal lobes.** *Brain.* 1985; **18**(4).
- Caffarra P, Vezzadini F, Dieci F, *et al.*: **Rey-Osterrieth complex figure: normative values in an Italian population sample.** *Neurol Sci.* 2002; **22**(6): 443–447. [PubMed Abstract](#) | [Publisher Full Text](#)
- Capitani E: **Normative data and neuropsychological assessment. Common problems in clinical practice and research.** *Neuropsychol Rehabil.* 1997; **7**(4): 295–310. [Publisher Full Text](#)
- Capitani E, Laiacona M: **Composite neuropsychological batteries and demographic correction: standardization based on equivalent scores, with a review of published data. The Italian Group for the Neuropsychological Study of Ageing.** *J Clin Exp Neuropsychol.* 1997; **19**(6): 795–809. [PubMed Abstract](#) | [Publisher Full Text](#)
- Cicerone KD, De Luca J: **Neuropsychological predictors of head injury rehabilitation outcome [abstract].** *J Clin Exp Neuropsychol.* 1990; **12**: 92.
- Cicerone KD, Wood JC: **Planning disorder after closed head injury: a case**

- study. *Arch Phys Med Rehabil.* 1987; **68**(2): 111–115.  
[PubMed Abstract](#)
11. Gazzaniga MS, Ivry RB, Mangun GR: **Cognitive Neuroscience.** New York: Norton & Company, Inc. 2002.
  12. Koss E, Patterson MB, Mack JL, *et al.*: **Reliability and validity of the Tinkertoy Test in evaluating individuals with Alzheimer's disease.** *Clin Neuropsychol.* 1998; **12**(3): 325–329.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  13. Krause M, Mahant N, Kotschet K, *et al.*: **Dysexecutive behaviour following deep brain lesions--a different type of disconnection syndrome?** *Cortex.* 2012; **48**(1): 97–119.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  14. Laiacona M, Inzaghi MG, De Tanti A, *et al.*: **Wisconsin Card Sorting Test: a new global score, with Italian norms, and its relationship with the Weigl sorting Test.** *Neurol Sci.* 2000; **21**(5): 279–291.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  15. Lezak MD: **The problem of assessing executive functions.** *Int J Psychol.* 1982; **17**(1–4): 281–297.  
[Publisher Full Text](#)
  16. Lezak MD, Howieson DB, Loring DV: **Neuropsychological Assessment (4<sup>th</sup> ed).** Oxford: Oxford University Press, 2004.
  17. Mendez MF, Ashla-Mendez M: **Differences between multi-infarct dementia and Alzheimer's disease on unstructured neuropsychological tasks.** *J Clin Exp Neuropsychol.* 1991; **13**(6): 923–932.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  18. Meyer JS: **Outer and Inner Confidence Intervals for Finite Population Quantile Intervals.** *J Am Stat Assoc.* 1987; **82**(397): 201–204.  
[Publisher Full Text](#)
  19. Mitrushina MN, Boone KB, Razani LJ, *et al.*: **Handbook of Normative Data for Neuropsychological Assessment (2<sup>nd</sup> ed).** Oxford: Oxford University Press, 2005.  
[Reference Source](#)
  20. Ownsworth T, Shum D: **Relationship between executive functions and productivity outcomes following stroke.** *Disabil Rehabil.* 2008; **30**(7): 531–540.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  21. Rizzo S, Venneri A, Papagno C: **Famous face recognition and naming test: a normative study.** *Neurol Sci.* 2002; **23**(4): 153–159.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  22. Spinnler H, Tognoni G (eds): **Standardizzazione e Taratura Italiana di Test Neuropsicologici.** *Ital J Neurol Sci.* 1987.  
[Reference Source](#)
  23. Stuss DT, Alexander MP: **Executive functions and the frontal lobes: a conceptual view.** *Psychol Res.* 2000; **63**(3–4): 289–298.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  24. Stuss DT, Levine B: **Adult clinical neuropsychology: lessons from studies of the frontal lobes.** *Annu Rev Psychol.* 2002; **53**: 401–433.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  25. Stuss DT: **Traumatic brain injury: relation to executive dysfunction and the frontal lobes.** *Curr Opin Neurol.* 2011; **24**(6): 584–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  26. Wilks SS: **Determination of sample sizes for setting tolerance limits.** *Ann Math Statist.* 1941; **12**(1): 91–96.  
[Publisher Full Text](#)
  27. Young DS: **Tolerance: an R package for estimating tolerance intervals.** *J Stat Softw.* 2010; **36**(5): 1–39.  
[Publisher Full Text](#)
  28. Crippa F, Cesana L, Daini R: **Dataset 1 in: Normative data for Lezak's Tinkertoy test in healthy Italian adults.** *F1000Research.* 2016.  
[Data Source](#)

# Open Peer Review

Current Referee Status: ? ?

---

## Version 1

Referee Report 03 October 2016

doi:[10.5256/f1000research.9051.r16486](https://doi.org/10.5256/f1000research.9051.r16486)



### Andrea Peru

Department of Neuroscience, Psychology, Drug Research, Child Health, University of Florence, Florence, Italy

This study provides normative data on the Lezak's Tinkertoy test for Italian population ranging in age from adolescence to older adulthood (range: 15-86 years).

In this vein, the word "adults" in the title seems to be not completely appropriate. Even more importantly, I don't see any reason why ethnicity should impact performance on this test. To sum up, I strongly suggest a shorter and more appealing title: Normative data for Lezak's Tinkertoy test.

Actually, so far normative data for this test are limited and based on a relatively small sample. Thus, I commend the authors for their intent to address this lack of evidence. There are, however, some shortcomings that should be addressed in a new version of the manuscript.

If my understanding is correct, the authors treated age as a continuous rather than discrete variable. If so, Table 1 is not necessary since it leaves the reader the idea that the age effect is spurious, due to the fact that the different age groups are not the same size. Alternatively, the authors could consider the opportunity to use statistical analysis like ANOVA – quite robust against different group sizes - rather than linear regression.

The main problem with the present paper, however, has to do with the fact that the authors missed a great opportunity to demonstrate that the Tinkertoy test has good construct validity. Previous studies reported that among elderly, demented, and traumatic brain injury patients the Tinkertoy test score has a significant, positive correlation with performances on the Trail Making and Wisconsin Card Sorting Test. In particular, the Tinkertoy complexity score turned out to be very sensitive to disorders of executive functions. In no way, however, can this be taken as convincing evidence that the Tinkertoy test is a reliable and valid instrument to assess executive functions in healthy people. I encourage the authors to provide further data on this topic.

Finally, I want to focus authors' attention on some grammatical and lexical issues.

As to grammar: Methods section, second paragraph: subject and verb are not in agreement as to number "Each subject..... were told...".

As to lexicon: For more than 100 years, in the field of experimental psychology, the term "subjects" has been used to describe people who take part in research and its use is still widely accepted. In the last decades, however, several psychological societies argued that the term "subject" is disrespectful, and recommended to replace it by "participants". Authors could consider this possibility.



Analogously, notwithstanding the taxonomic label “frontal lobe syndrome” is still very popular among clinicians, it has really had its day and should be replaced by “dysexecutive syndrome” or “prefrontal lobe dysfunction” according to whether the emphasis is put on the function or the localization.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

**Competing Interests:** No competing interests were disclosed.

Referee Report 13 June 2016

doi:[10.5256/f1000research.9051.r14328](https://doi.org/10.5256/f1000research.9051.r14328)



**Luigi Trojano**

Department of Psychology, Second University of Naples, Caserta, Italy

The authors present a normative study for an unstructured test aimed at assessing cognitive flexibility and planning abilities. The methods adopted to obtain normative values are often used in Italy and allow comparing individuals’ performance on neuropsychological tests tapping different cognitive domains. The study sample is similar to that enrolled in other normative studies.

However, the authors also enrolled 29 late adolescents (15-18 year-old) in their sample, and this is not usual in normative studies on neuropsychological tests. I suggest deleting these individuals from the sample, as development and maturation of executive processes across adolescence is quite variable and should be addressed by dedicated normative studies.

The authors should also consider several limitations of their study, and acknowledge them in their final remarks. For instance, the authors stated in their discussion that they enrolled a “wide” sample, “representative of the Italian population”. I believe that the authors should provide more details about procedures and methods adopted for recruitment of participants, and likely tone down claims about the extension of the sample and its representativeness of the “Italian population”. Moreover, the study did not provide data about psychometric properties of the test such as convergent or divergent validity, test-retest reliability, or inter-rater agreement. I understand that to address some of these properties is outside the authors’ scope, but I believe that at least inter-rater agreement is important to reassure the reader about clinical applicability of the present test.

Minor points: the authors should provide more details about the exact instructions for examinees, about time limits, and about the pieces to be used, also to make clear for the reader which types of errors are considered as “misfit” or “incomplete fit”. Info about availability of test material might be useful.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

**Competing Interests:** No competing interests were disclosed.