Tower of London test — short version

Pavel Harsa[1], Michaela Břeňová[1], Ondřej Bezdíček[2], Lucie Kališová[1],
Jana Heidingerová[1], Kateřina Schönová[1], Jiří Michalec[1]

[1]Department of Psychiatry, First Faculty of Medicine and General University Hospital,
Charles University Kateřinská, Prague, Czech Republic
[2]Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine and General University Hospital,
Charles University Kateřinská, Prague, Czech Republic

ABSTRACT

Tower of London (ToL) is a neuropsychological method for assessing planning ability. In this study, we attempted to introduce a shorter version of ToL. Two studies were carried out. In Study 1, patients with mild cognitive impairment due to Parkinson's disease (n = 36) and a control group (n = 225) were included in order to select a suitable short version of ToL for Study 2. In Study 2, patients with schizophrenia (n = 30) and a control group (n = 31) were included in order to assess psychometric properties of the shortened version of ToL. Based on psychometric evaluations in Study 1, we offered three possible shortened versions. In Study 2, all three shortened versions proved to have good discriminative validity in our schizophrenia sample, but interestingly not in the healthy sample. We concluded that the use of shortened ToL is possible in specific psychiatric/neurological patient groups, although we do not recommend doing so in healthy individuals.

Key words: Tower of London, short version, executive functions, planning ability, problem solving, neuropsychological assessment

(*Neurol Neurochir Pol 2022; 56 (4): 365–370*)

Introduction

'Executive functions' is an umbrella term for various cognitive processes including problem solving, planning, organisational skills and inhibitory control. In neuropsychology, information about patients' planning ability, problem solving etc. can be obtained using multiple methods. These might include neuropsychological tests, observations, interviews, questionnaires, and rating scales. For the purposes of assessing executive functions, and particularly planning ability, Hunter and Sparrow [1] mention not only Tower of London tests, but also maze tasks, several variants of the categorisation tests (e.g. Wisconsin Card Sorting Test), copying designs in the Rey-Osterrieth Complex Figure (ROCF), and verbal fluency tasks.

The Tower of London (ToL) task was first proposed by Shallice [2] as an alternative to the Tower of Hanoi. Nowadays, the ToL is frequently used in clinical psychology settings for an assessment of executive functioning, namely planning ability [3, 4].

Specific clinical populations have been proven to have poor performance when it comes to the ToL task, especially patients with frontal lobe impairment [2, 4, 5]. The ToL is also being used in assessing patients with Parkinson's Disease [6–8], Huntington's Disease [9–11], schizophrenia [8, 12, 13], dementia or mild cognitive impairment [14–17], and autism spectrum disorder [18–20].

Over the years, various versions of the original ToL have been introduced [18, 21–23], using an inconsistent variety of problems, performance measures and methodologies [3].

For example, a standardised version of the ToL test has been developed for use as a part of the Delis–Kaplan Executive Function System [24]. The Tower test in D-KEFS is longer than the original ToL, but covers a wider spectrum of performance, from severe deficits to superior performance. In the D-KEFS Tower Test, there are nine different towers to be completed with various levels of difficulty. Testing begins with simple towers requiring only 1–3 moves, and gradually becomes more difficult with towers requiring up to 26 moves [25].

Another example is the Tower of Toronto, designed specifically for patients with striatal disorders affecting procedural learning [26].

The original Shallice version implements a block base with three rods of different lengths, and three balls of different colours (red, green and blue). The three balls are placed on the rods in a starting position and for each problem the balls have to be moved from the starting position into the target position while respecting a set of rules and a maximum number of moves [23].

A standardised Czech version of ToL was however lacking. In recent years, Michalec et al. [27] published a standardised Czech version of Shallice's original ToL and provided preliminary normative data for healthy elderly people. Later, Michalec et al. [8] published definitive normative data for the whole Czech population, and compared various scoring systems suitable for the original Shallice version of ToL. The subsequent release of the Czech manual and props for standardised administration of ToL Shallice version [28] made the method more available in the Czech Republic.

Raizner et al. [29] closely examined an alleged 'ceiling effect' of ToL when testing adolescents and young adults. According to this study, ToL consistently shows satisfactory sensitivity to the development of planning abilities in children; however, sensitivity to the development of planning ability in healthy adolescents and young adults is inconsistent. The authors of the study further explored whether a ceiling effect was to blame, and they introduced an extended version of ToL (ToL-E). The ToL-E included an additional eight problems with same rules and props as the original Shallice ToL, but they required 6, 7 or 8 moves to complete, as opposed to the maximum 5 moves in the hardest task of the original ToL.

Bearing in mind this extended version of the ToL, we considered it important to introduce a shortened version of the method, especially for an adult clinical population. In differentiating between various levels of executive functioning in young adults, we might find using more tasks useful [29], but in screening for severely cognitively impaired adults in a clinical environment, we felt this might be better accomplished in fewer tasks.

To test this theory, we introduce our current study of the brief screening version of ToL using only three chosen ToL tasks instead of the usual number of 12.

Test shortening is a well observed trend in clinical psychology, most often in order to decrease testing time and to ensure patient cooperation [30]. However, several challenges arise when shortening an established psychological method, as we go against the 'classical psychometric assumption' in which many items are essential to obtain valid and reliable measures [31, 32]. We chose to deal with this issue by choosing a statistics-driven strategy, which we explain below. This research was approved by the ethics committee of the First Faculty of Medicine, Charles University, Prague, Czech Republic and the authors did not have any financial interest in the research.

Study 1

The aim of Study 1 was to select for further psychometric evaluation (Study 2) the optimal short version of ToL from all 4,095 possible combinations. Primarily, we wanted to achieve a discriminative validity of the chosen combinations as close to the discriminative validity of all the 12 original items combined.

Methods

Procedure

Both groups, i.e., patients with mild cognitive impairment (MCI) in Parkinson's Disease (PD) and the healthy control group, were administered the original 12-item Shallice version of ToL. Their performance was evaluated based on the Shallice original scoring system: 3 points for successful task completion within 15 seconds, 2 points for completion within 30 seconds, 1 point for completion within 60 seconds, and 0 points for completion in more than 60 seconds. The number of trials was irrelevant, but the time in each trial was added to the total time sum. For a more detailed explanation of the scoring system, see Michalec et al. [8].

The PD-MCI group was also administered a complex neuropsychological test battery according to level II criteria for PD-MCI as established by Litvan et al. [33]. The control group was administered DRS-2 (Dementia Rating Scale 2) [34, 35] as a screening for possible cognitive impairment.

Participants

As a suitable model of executive functions deficit (including planning ability), we chose to study a sample of patients with mild cognitive impairment (MCI) in Parkinson's Disease (PD). Participants were recruited at the Department of Neurology at the First Faculty of Medicine and General University Hospital in Prague, Czech Republic. The total sample of n = 46 patients was 80% male, with average years of education of 13.78 (± 2.66) and an average age of 59.52 (± 7.29). The diagnosis of PD-MCI in our sample was established based on a complex neuropsychological test battery in accordance with contemporary diagnostic criteria on level II [33].

The model of intact planning ability was represented by a sample of n = 225 healthy volunteers (control group) without any disease with a possible impact on their cognitive functioning. Also, all the volunteers in our sample had to fulfil test criteria of a DRS-2 (Dementia Rating Scale 2) score > 141 [30, 31]. Fifty per cent of this sample were male, the average years of education was 14.16 (± 2.95), and average age was 54.57 (± 13.28).

Table 1. ToL test items included in shortened versions

	ToL 01	ToL 02	ToL 03	ToL 04	ToL 05	ToL 06	ToL 07	ToL 08	ToL 09	ToL 10	ToL 11	ToL 12
ToL-four items						x		x	x	x		
ToL-five items						x		x	x	x	x	
ToL-six items						x		x	x	x	x	x

Statistical analysis

Receiver operating characteristic (ROC) analysis was carried out, and area under curve (AUC) was determined for all 4,095 combinations.

Results

206 of 4,095 possible combinations had a similar AUC to that of all 12 items combined (AUC = 0.753). However, none of the combinations had significantly better clinical discriminative validity. AUC ranged from 0.753 to 0.769 (see Appx). Out of these 206 combinations of items, only four were four-item combinations. Out of these four-item combinations, all consisted of four and five moves items. None of them included less than four moves items. Taking this into account, for further evaluation we chose only combinations of items with four moves or more, which rather narrowed our selection. The chosen items were items 6-8-9-10 for a four-item version, items 6-8-9-10-11 for a five-item version, and items 6-8-9-10-11-12 for a six-item version. This particular selection enabled parallel administration of all three versions chosen for the purpose of their psychometric evaluation and comparison (Study 2), i.e., to. administer once a six-item version and only to subtract items for psychometric evaluation of a given short version with a minimal risk of influencing the results via a learning effect or via the order of items.

Study 2

The aim of Study 2 was to obtain a research sample for assessing basic psychometric properties of the shortened version of ToL.

Methods

Procedure

Based on the results of Study 1, from 206 possible combinations, three versions of shortened ToL were chosen: one with four items, one with five items, and one with six items. Items included in a given version of shortened ToL are set out in Table 1. Schizophrenia patients and controls were administered the ToL-six-item version. Shorter versions were subtracted from the ToL-six-item version. There was no possible interference by different learning curves etc. The scoring system used was the same as in the original version.

Participants

The research sample consisted of n = 30 patients with schizophrenia from the Department of Psychiatry of the First Faculty of Medicine and General University Hospital in Prague. The sample was 77% male, with average years of education of 13.36 (± 3.12) and average age of 31.37 (± 6.27). All patients were in remission at the time of assessment.

The control group consisted of n = 31 socio-demographically paired healthy volunteers without any history of a disease with a possible effect on their cognitive functioning, and who also fulfilled the test criteria of DRS-2 score > 141 [30, 31]. The sample was 74% male, with average years of education of 15.53 (± 3.37) and average age of 33.42 (± 3.81).

Participants were excluded from the study if they were found to have a history of traumatic injury of the central nervous system, a neurological disorder, a premorbid intellectual disability, or any substance dependence (with the exception of nicotine).

Statistical analysis

Receiver operating characteristic (ROC) analysis was carried out for each short ToL version. Area under curve (AUC) was computed and also values of sensitivity and specificity for given cut-off scores. Descriptive statistics for raw ToL scores were also computed.

Results

Table 2 shows descriptive statistics of ToL raw scores of each short ToL version.

From the results listed in Table 3, we can deduce that the discriminative validity (meaning AUC values) of individual shortened versions of ToL was equal and clinically valuable. The reliability of shortened versions of ToL was analyzed by using internal consistency (Cronbach's alpha) separately for both the schizophrenia and control groups. The results yielded insufficient Cronbach's alpha values for the control group, yet conversely satisfactory results when it came to the schizophrenia group. The ToL-four-item version had slightly lower internal consistency than the ToL-five-item and ToL-six-item versions.

Table 4 sets out the values of sensitivity and specificity for given cut-off scores for each of the ToL short versions, showing in more detail their discriminative validity.

Table 2. Descriptive statistics of ToL short versions raw scores

	CG		SCH	
	M ± SD	(MIN-MAX)	M ± SD	(MIN-MAX)
ToL-four items	8.81 ± 1.89	(5–12)	5.6 ± 2.67	(0–10)
ToL-five items	10.74 ± 2.44	(6–15)	7.07 ± 3.39	(0–13)
ToL-six items	12.03 ± 2.74	(7–17)	8.1 ± 3.91	(0–15)

CG — control group; M ± SD — mean ± standard deviation; MIN-MAX — minimum raw ToL score minus maximum raw ToL score; SCH — patients with schizophrenia
For ToL-four items, minimal possible raw score is 0 and maximum is 12
For ToL-five items, minimal possible raw score is 0 and maximum is 15
For ToL-five items, minimal possible raw score is 0 and maximum is 18

Table 3. Results — Discriminative validity of shortened versions of ToL

	ToL 06	ToL 08	ToL 09	ToL 10	ToL 11	ToL 12	AUC	95% CI AUC	Cronbach alpha	
									NC	SCH
ToL-four items	x	x	x	x			0.833	0.732–0.935	0.22	0.64
ToL-five items	x	x	x	x	x		0.804	0.695–0.914	0.40	0.73
ToL-six items	x	x	x	x	x	x	0.788	0.675–0.900	0.45	0.76

Table 4. Cut-off scores for ToL short versions and their sensitivity and specificity

ToL-four items			ToL-five items			ToL-six items		
Cut-off	SE	SP	Cut-off	SE	SP	Cut-off	SE	SP
4/5	33	100	4/5	20	100	4/5	17	100
5/6	53	97	5/6	30	100	5/6	23	100
6/7	63	90	6/7	47	94	6/7	43	100
7/8	77	71	7/8	53	90	7/8	47	94
8/9	83	55	8/9	57	74	8/9	50	90
9/10	90	39	9/10	80	71	9/10	57	77
10/11	100	19	10/11	87	61	10/11	67	71
11/12	100	10	11/12	90	45	11/12	77	58
			12/13	93	23	12/13	90	48

Discussion

The main aim of this study was to examine a shorter version of ToL for the clinical purposes of assessing executive functioning (planning ability) of psychiatric patients. As already mentioned, there has been a lack of similar efforts to establish shortened versions of ToL in the scientific community. In our clinical experience, the original Shallice ToL version [2] can prove difficult and time consuming when administered to patients with severe psychiatric or neurological diseases.

Our study shows that shortening ToL is possible. In our sample, we used four-, five- and six-item versions. The shortened versions yielded a satisfactory value of internal consistency when it came to the schizophrenia sample, but interestingly not when it came to the control group. Interpreting these findings, we hypothesise that the use of a shortened version

is indeed appropriate only in patient samples with expected problems in the executive area. However, for assessing executive functions in healthy individuals, the use of short versions of ToL is not recommended.

The implications of test shortening can be several. The original psychometric properties of the method can be altered and need to be established again for the shortened version. As already mentioned, shortening of neuropsychological methods goes against the 'traditional' psychometric idea of many items being essential for obtaining valid and reliable measures [31, 32]. That is why all endeavours towards establishing such shortened versions must be statistically-driven and thoroughly evaluated.

There are several limitations to our study. The samples used were rather small in size and not socio-demographically well balanced when it came to clinical groups. However, this is often an inevitable issue with patient groups with a specific diagnosis

— each diagnostic group has its own specific socio-demographic characteristics, which is then naturally reflected in the patient sample. Furthermore, our study participants were mostly chosen opportunistically and based on certain criteria, rather than via random sampling. Also, the psychometric properties were derived from a sample of patients with schizophrenia. The study does not provide comparisons with other clinical groups, for example neurological patients. Using the ToL short version may be more challenging for neurological patients who have initiation problems, i.e., find it hard to start performing new tasks, and for patients with procedural memory impairment. We suggest that these issues and other related questions concerning the appropriateness of this test for diverse neurological populations should be the subject of further research.

Conclusions

Our main conclusion, as well as recommendation, is that using the ToL short version can possibly prove sufficient in certain clinical samples, and we encourage further research efforts in that area.

Regarding the equal discriminative validity of all three ToL short versions, and the fact that the ToL-four item version had a slightly lower internal consistency in the patient group, we recommend the ToL-five item version for clinical use. We recommend the ToL-five rather than the ToL-six simply because it is shorter. Cut-off scores for preliminary clinical use were presented, although this cannot replace standardisation including normative data.

References

1. Hunter S, Sparrow E. Executive Function and Dysfunction: Identification, Assessment and Treatment. Cambridge University Press, Cambridge 2012.
2. Shallice T. Specific impairments of planning. Philos Trans R Soc Lond B Biol Sci. 1982; 298(1089): 199–209, doi: 10.1098/rstb.1982.0082, indexed in Pubmed: 6125971.
3. Keith Berg W, Byrd D. The Tower of London spatial problem-solving task: enhancing clinical and research implementation. J Clin Exp Neuropsychol. 2002; 24(5): 586–604, doi: 10.1076/jcen.24.5.586.1006, indexed in Pubmed: 12187443.
4. Newman SD, Pittman G. The Tower of London: a study of the effect of problem structure on planning. J Clin Exp Neuropsychol. 2007; 29(3): 333–342, doi: 10.1080/13803390701249051, indexed in Pubmed: 17454353.
5. Sullivan JR, Riccio CA, Castillo CL. Concurrent validity of the tower tasks as measures of executive function in adults: a meta-analysis. Appl Neuropsychol. 2009; 16(1): 62–75, doi: 10.1080/09084280802644243, indexed in Pubmed: 19205950.
6. Culbertson WC, Moberg PJ, Duda JE, et al. Assessing the executive function deficits of patients with Parkinson's disease: utility of the Tower of London-Drexel. Assessment. 2004; 11(1): 27–39, doi: 10.1177/1073191103258590, indexed in Pubmed: 14994951.
7. Köstering L, Schmidt CSM, Egger K, et al. Assessment of planning performance in clinical samples: Reliability and validity of the Tower of London task (TOL-F). Neuropsychologia. 2015; 75: 646–655, doi: 10.1016/j.neuropsychologia.2015.07.017, indexed in Pubmed: 26197091.
8. Michalec J, Bezdicek O, Nikolai T, et al. A comparative study of tower of london scoring systems and normative data. Arch Clin Neuropsychol. 2017; 32(3): 328–338, doi: 10.1093/arclin/acw111, indexed in Pubmed: 28096070.
9. Ho AK, Sahakian BJ, Brown RG, et al. NEST-HD Consortium. Profile of cognitive progression in early Huntington's disease. Neurology. 2003; 61(12): 1702–1706, doi: 10.1212/01.wnl.0000098878.47789.bd, indexed in Pubmed: 14694033.
10. Mörkl S, Müller NJ, Blesl C, et al. Problem solving, impulse control and planning in patients with early- and late-stage Huntington's disease. Eur Arch Psychiatry Clin Neurosci. 2016; 266(7): 663–671, doi: 10.1007/s00406-016-0707-4, indexed in Pubmed: 27372072.
11. Watkins LH, Rogers RD, Lawrence AD, et al. Impaired planning but intact decision making in early Huntington's disease: implications for specific fronto-striatal pathology. Neuropsychologia. 2000; 38(8): 1112–1125, doi: 10.1016/s0028-3932(00)00028-2, indexed in Pubmed: 10838146.
12. Grover S, Sahoo S, Nehra R, et al. Association of neurocognitive deficits and insight in schizophrenia. Asian J Psychiatr. 2018; 36: 112–117, doi: 10.1016/j.ajp.2018.07.001 , indexed in Pubmed: 30064045.
13. Zhu Ye, Liu X, Wang H, et al. Reduced prefrontal activation during Tower of London in first-episode schizophrenia: a multi-channel near-infrared spectroscopy study. Neurosci Lett. 2010; 478(3): 136–140, doi: 10.1016/j.neulet.2010.05.003, indexed in Pubmed: 20457217.
14. de Paula JJ, Moreira L, Nicolato R, et al. The Tower of London Test: different scoring criteria for diagnosing Alzheimer's disease and mild cognitive impairment. Psychol Rep. 2012; 110(2): 477–488, doi: 10.2466/03.10.13.PR0.110.2.477-488, indexed in Pubmed: 22662402.
15. Franceschi M, Caffarra P, Savarè R, et al. Tower of London Test: A Comparison between Conventional Statistic Approach and Modelling Based on Artificial Neural Network in Differentiating Fronto-Temporal Dementia from Alzheimer's Disease. Behavioural Neurology. 2011; 24(2): 149–158, doi: 10.1155/2011/754659.
16. Marchegiani A, Giannelli MV, Odetti PR. The tower of London test: a test for dementia. Aging Ment Health. 2010; 14(2): 155–158, doi: 10.1080/13607860903228804, indexed in Pubmed: 19937502.
17. Rainville C, Amieva H, Lafont S, et al. Executive function deficits in patients with dementia of the Alzheimer's type A study with a Tower of London task. Archives of Clinical Neuropsychology. 2002; 17(6): 513–530, doi: 10.1093/arclin/17.6.513.
18. Culbertson WC, Zillmer EA. The Tower of London(DX): a standardized approach to assessing executive functioning in children. Arch Clin Neuropsychol. 1998; 13(3): 285–301, indexed in Pubmed: 14590643.
19. Just MA, Cherkassky VL, Keller TA, et al. Functional and anatomical cortical underconnectivity in autism: evidence from an FMRI study of an executive function task and corpus callosum morphometry. Cereb Cortex. 2007; 17(4): 951–961, doi: 10.1093/cercor/bhl006, indexed in Pubmed: 16772313.
20. Unterrainer JM, Rahm B, Loosli SV, et al. Psychometric analyses of the Tower of London planning task reveal high reliability and feasibility in typically developing children and child patients with ASD and ADHD.

Child Neuropsychol J Norm Abnorm Dev Child Adolesc. 2020 Feb.; 26(2): 257–273.

21. Köstering L, Nitschke K, Schumacher FK, et al. Test-retest reliability of the Tower of London Planning Task (TOL-F). Psychol Assess. 2015 Sep.; 27(3): 925–931.

22. Tunstall JR, O'Gorman JG, Shum DHK. A four-disc version of the Tower of London for clinical use. J Neuropsychol. 2016; 10(1): 116–129, doi: 10.1111/jnp.12060, indexed in Pubmed: 25420463.

23. Unterrainer JM, Rahm B, Halsband U, et al. What is in a name: comparing the Tower of London with one of its variants. Brain Res Cogn Brain Res. 2005; 23(2-3): 418–428, doi: 10.1016/j.cogbrainres.2004.11.013, indexed in Pubmed: 15820648.

24. Delis D, Kaplan E, Kramer J. Delis-Kaplan Executive Function System. PsycTESTS Dataset. 2012, doi: 10.1037/t15082-000.

25. Yochim BP, Baldo JV, Kane KD, et al. D-KEFS Tower Test performance in patients with lateral prefrontal cortex lesions: the importance of error monitoring. J Clin Exp Neuropsychol. 2009; 31(6): 658–663, doi: 10.1080/13803390802448669, indexed in Pubmed: 19031323.

26. Saint-Cyr JA, Taylor AE, Lang AE. Procedural learning and neostriatal dysfunction in man. Brain J Neurol. Brain. 1988; 111(Pt 4): 941–959, doi: 10.1093/brain/111.4.941 , indexed in Pubmed: 2969762.

27. Michalec J, Bezdíček O, Nikolai T, et al. Standardization of the Czech Version of the Tower of London Test – Administration, Scoring, Validity. Česká a slovenská neurologie a neurochirurgie. 2014; 77/110(5): 596–601, doi: 10.14735/amcsnn2014596.

28. Bezdicek O, Michalec J, Shallice T. Londýnská věž (ToL), Schalliceova verze ToL. Diagnostický test. Univerzita Karlova, Prague 2018.

29. Raizner RD. Raising the ceiling: the Tower of London--extended version. Dev Neuropsychol. 2002; 21(1): 1–14, doi: 10.1207/S15326942DN2101_1, indexed in Pubmed: 12058833.

30. Kleka P, Paluchowski W. Shortening of psychological tests – assumptions, methods and doubts. Polish Psychological Bulletin. 2017; 48(4): 516–522, doi: 10.1515/ppb-2017-0058.

31. Anastasi A. Psychological testing. Macmillan/Collier Macmillan, London/New York 1982.

32. Kruyen P, Emons W, Sijtsma K. On the Shortcomings of Shortened Tests: A Literature Review. International Journal of Testing. 2013; 13(3): 223–248, doi: 10.1080/15305058.2012.703734.

33. Litvan I, Goldman J, Tröster A, et al. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: *Movement* Disorder Society Task Force guidelines. Movement Disorders. 2012; 27(3): 349–356, doi: 10.1002/mds.24893.

34. Bezdicek O, Michalec J, Nikolai T, et al. Clinical validity of the mattis dementia rating scale in differentiating mild cognitive impairment in parkinson's disease and normative data. Dement Geriatr Cogn Disord. 2015; 39(5-6): 303–311, doi: 10.1159/000375365, indexed in Pubmed: 25792240.

35. Kreutzer JS, Leitten C, Mattis S. Mattis Dementia Rating Scale. In: DeLuca J, Caplan B. ed. Encyclopedia of Clinical Neuropsychology. Springer, New York 2011.