Kastamonu Education Journal, 2021, Vol. 29, No:3, 705-719
doi: 10.24106/kefdergi.790903

|Araştırma Makalesi / Research Article|

Investigation of Planning Skills in Children and Adults with the Computerized 4-Disc Version of The Tower of London Test

4 Diskli Londra Kulesi Testinin Bilgisayarlı Versiyonu ile Çocuklarda ve Yetişkinlerde Planlama Becerisinin İncelenmesi

Elif GÜVEN DEMİR[1], **Yücel ÖKSÜZ**[2]

Keywords

1.tower of london test
2.planning
3.children
4.adult
5.computer

Anahtar Kelimeler

1.londra kulesi testi
2.planlama
3.çocuk
4.yetişkin
5.bilgisayar

Başvuru Tarihi/Received
05.09.2020

Kabul Tarihi /Accepted
21.04.2021

Abstract

Purpose: Planning ability is one of the core abilities and plays an essential role in problem-solving and decision-making processes. It is mainly critical in childhood to manage school-related demands and carry out daily activities at older ages. This study aimed to develop and apply a computerized version of the 4-disc Tower of London (TOL) test for planning skills.

Design: The research was conducted by three studies, including children, adults, and combined samples. The research consisted of children and adults who were 5-53 years old. Nonparametric tests were used, thereby that the research data had a non-normally distributed pattern.

Findings: Research results showed significant relationships between age and planning ability, the number of problems solved on the first trial, and the number of problems solved within three trials. Results showed that increasing planning ability and the number of problems solved on the first trial was associated with increasing age. The increasing number of problems solved within three trials was associated with decreasing age. There was no significant relationship between age and planning time.

Highlights: Research results showed that gender has an effect on planning time in the adult group. Results showed that TOL has an average medium difficulty level and good item discrimination level in the children sample. Besides, item difficulty and discrimination levels of TOL in the adult sample were acceptable. Results showed that the internal consistency level of TOL was acceptable. In this study, only item difficulty, discrimination, and internal consistency analysis were performed, and studies including detailed validity and reliability analysis can be conducted in future studies.

Öz

Çalışmanın Amacı: Planlama becerisi, problem çözme ve karar verme sürecinde önemli bir rol oynayan temel becerilerden biridir. Özellikle çocukluk döneminde okulla ilgili öğrenme süreçlerinin yönetilmesinde, ilerleyen yaşlarda ise günlük hayat aktivitelerinin yerine getirilmesinde önemli bir yere sahiptir. Bu çalışmanın amacı, planlama becerisini ölçen 4 diskli Londra Kulesi (TOL) testinin bilgisayarlı bir versiyonunu geliştirmek ve uygulamaktır.

Materyal ve Yöntem: Araştırma, çocukları, yetişkinleri ve her ikisini de içeren grupların yer aldığı üç aşamada gerçekleştirilmiştir. Araştırma, 5-53 yaş arası çocuk ve yetişkin bireylerin katılımıyla gerçekleştirilmiştir. Araştırma verilerinin normal dağılım göstermemesi sebebiyle verilerin analizinde parametrik olmayan testler kullanılmıştır.

Bulgular: Araştırma sonuçları, yaş ile planlama becerisi, ilk denemede çözülen soru sayısı, üç denemede çözülen soru sayısı arasında anlamlı ilişkiler olduğunu göstermektedir. Sonuçlar, artan planlama becerisinin ve ilk denemede çözülen sorun sayısının artan yaşla ilişkili olduğunu göstermektedir. Üç denemede çözülen soru sayısının yaşın azalmasıyla ilişkili olduğu, yaş ile planlama zamanı arasında anlamlı bir ilişkinin olmadığı tespit edilmiştir.

Önemli Vurgular: Araştırma sonuçları, yetişkin grupta cinsiyet değişkeninin zamanı planlamada etkili olduğunu göstermektedir. Sonuçlar TOL'un çocuk örnekleminde ortalama düzeyde zorluk ve iyi düzeyde de ayırıcılık düzeyine sahip olduğunu göstermektedir. Yetişkin örnekleminde ise TOL kabul edilebilir düzeyde ayırıcılık ve zorluk düzeyine sahip olduğu tespit edilmiştir. Sonuçlar TOL'un iç tutarlılık düzeyinin kabul edilebilir düzeyde olduğunu göstermektedir. Bu araştırmada yalnızca madde güçlük, ayırıcılık ve iç tutarlılık analizi gerçekleştirilmiş olup, ilerleyen araştırmalarda ayrıntılı geçerlik ve güvenirlik çalışmaları yapılabilir.

[1] **Corresponded Author,** Düzce University, Faculty of Education, Primary Education Department, Düzce, TURKEY; https://orcid.org/0000-0001-6685-5341
[2] Ondokuz Mayıs University, Faculty of Education, Department of Educational Sciences, Samsun,TURKEY; https://orcid.org/ 0000-0002-9310-7506

Citation: Güven Demir, E., & Öksüz, Y. (2021). Investigation of planning skills in children and adults with the computerized 4-disc version of the tower of london test. *Kastamonu Education Journal, 29*(3), 705-719. doi: 10.24106/kefdergi.790903

INTRODUCTION

Planning ability is an essential cognitive skill as a part of executive functions. Executive function is an umbrella term that refers to multiple processes and includes complex cognitive skills and their relationship to adjust target-oriented behavior. Hughes (2002) described executive functions as a complex cognitive process and flexible target-oriented behaviors, including planning, inhibition control, flexibility in attention, and working memory. In addition, there are particular core abilities; planning, focusing, and sustaining attention, temporal organization, executing target-oriented behaviors, blocking inappropriate responses, being fluent and flexible in thought and action, monitoring behaviors, and using feedback to regulate behavior (Tunstall, 1999; Fossati, Ergis & Allilaire, 2002).

There is a debate about whether planning ability is a specific executive function (Rönmark, 2014). Asato, Sweeney, & Luna (2006) suggested that planning is a complex cognitive process that requires integrating core cognitive components, including response inhibition and working memory. McCormack & Atance (2011) mentioned three claims about planning and executive function ability. The first claim suggests that planning develops as part of a general, unfractionated executive function ability. The second claim is that planning itself develops as a relatively separate sub-component of executive functions (Levin & Hanten, 2005; Welsh, Pennington, & Grossier, 1991; as cited McCormack & Atance, 2011). The third claim is that children's performance on planning tasks improves due to developmental changes in the subcomponents of executive functions (McCormack & Atance, 2011). The complex structure and subcomponents of executive functions may require planning ability in different ways. Rönmark (2014) suggested that planning is seen as a possible combination of different executive functions.

Planning is a complex form of action that consists of a consciously predetermined sequence of actions that will be adequate for achieving a task (Pea, 1982). Owen (1997) described planning ability as "thinking ahead," which is a central element of many aspects of complex behavior and is a basic necessity of many cognitive and motor tasks. At the core of diversified definitions of planning involves the mental representation and/or behavioral execution of actions to achieve a future goal (Tunstall, 1999). Harlow (1869) can be considered the ancestor of the concept of planning ability in the anatomic structure. Owen (1997) mentions that Harlow (1869) argued that frontal lobe lesions in humans result in a loss of "planning skill." Besides, Luria (1966) distinguished three functional units in the brain where each unit has the planning role. The prefrontal cortex has a central function in cognitive control in arranging thought and action in accordance with internal goals (Miller & Cohen, 2001). Dockery, Hueckel-Weng, Birbaumer & Plewnia (2009) indicated that planning abilities are preconditions for successful problem solving and effective behavior, and patients with frontal lobe pathology as schizophrenia, depression, and lesions, commonly show executive function impairment.

Diverse definitions of planning bring along the differentiation of naming and thoughts at the point of exit. Some researchers consider that planning occurs only before the commencement of the action, although most consider that planning occurs both prior to and during the action (Scholnick & Friedman, 1987, as cited by Tunstall, 1999). Pea (1982) mentioned that the planning process comprises four steps. The first step represents the planning problem situation, a task that requires the planner to define the seal state, define the problem state, note the differences between the problem and goal states, and determine planning constraints (Pea, 1982). The following steps are plan construction, which includes requiring the formulation of a plan to eliminate the differences between the problem and the goal state, plan execution, and planning process is remembering. Besides, planning was distinguished within three different levels: activity planning, action planning, and operation planning in activity theory (Leontjev 1978; as cited by Das & Georgiou, 2016). Grafman (1989) and Shalice (1982) suggested that planning can be described as a double-level process: (1) The formulation level relies on the ability to mentally develop a logical strategy to predetermine the course of action aimed at achieving a specific goal. (2) The execution level is concerned with the competence of monitoring and guiding the execution of the plan towards a successful conclusion (as cited in Allain et al., 2005).

The planning ability emerges early and has a long developmental course associated with the prefrontal cortex's development, which continues to develop through adolescence and early adulthood (Wilding, Munir, & Cornish, 2001 as cited in Less, 2008). Planning ability is widely regarded as an important developmental achievement (McCormack & Atance, 2011). Particularly in childhood, planning ability is critical for social and cognitive development, such as managing school-related demands and balancing various activities (Blair, 2002; as cited in Less, 2008). In addition, planning ability is a higher-level cognitive process and plays a vital role in problem-solving and decision making (Mahapatra, 2016).

The planning ability is involved in a range of important life skills such as cooking, shopping, and various occupational tasks (Phillips, Kliegel & Martin, 2006). Cahn-Weiner, Malloy, Boyle, Marran & Salloway (2000) claimed that planning ability as an executive function is a better predictor of the ability to carry out daily activities in old age than more computerized cognitive measures such as intelligence and memory (Phillips et al., 2006).

The two distinct approaches investigating the planning ability are observing the planning ability in the natural environment while performing daily life tasks and measuring the planning ability with the standard test in a clinical or laboratory environment (Tunstall, 1999). Both approaches have advantages and disadvantages in terms of investigating the planning process properly. Lack of control on variables that may affect the planning process in daily life and lack of contact with daily life tasks are disadvantages of each approach. However, the potential of controlling environmental and task variables provides an advantage to the standard tests to investigate daily life planning ability (Tunstall, 1999).

Tower task tests have been popular in measuring planning and problem-solving ability. The initial tower test, the Tower of Hanoi, was invented by the French mathematician Edouard Lucas and marketed as a toy in 1883 (Gardner, 1959; as cited by Rönnlund, Lövden & Nilsson, 2001). Simon (1975) used the Tower of Hanoi to measure problem-solving skills to show that even in simple problem environments, numerous distinct solution strategies are available, and different subjects may learn different strategies. The Tower of Hanoi consists of three pegs and several disks of varying sizes (Goel & Grafman, 1995). Given a start state, in which the disks are stacked on one or more pegs, the task is to reach a goal state in which the disks are stacked in descending order on a specified peg (Goel & Grafman, 1995). The Tower of Hanoi is traditionally administered so that the puzzles require an increasing number of moves for a solution, and the task is scored as the number of puzzles solved or the highest level (number of steps to solution) reached (Kopecky, Chang, Klorman, Thatcher, & Borgstedt, 2005). Shallice (1982) aimed to develop an instrument with gradual difficulty levels providing a greater variety of qualitatively different problems than the Tower of Hanoi (Unterrainer, Rahm, Halsband, & Kaller, 2005). Although the Tower of Hanoi test offers several advantages as it is portable, visually stimulating, less threatening than many problem-solving tests, and is easy to apply from early ages to old ages, there are doubts about its ability to measure planning (Tunstall,1999; Goel & Grafman, 1995). Shallice (1982) claimed that the Tower of Hanoi test did not have the potential to have several quantitatively different versions of comparable difficulty (Shallice & Burgess, 1991, as cited by Tunstall, 1999). Shallice (1982) has developed the Tower of London test (TOL) based on the Tower of Hanoi test in artificial intelligence studies as a planning task.

The Tower of London test has many variations. The original TOL has three discs, colored red, green, and blue, and three poles in increasing heights (see figure 1). Respondents must rearrange the discs to match a target arrangement and do so in a specified number of moves (Andrews, Halford, Chappell, Maujean, & Shum, 2014).

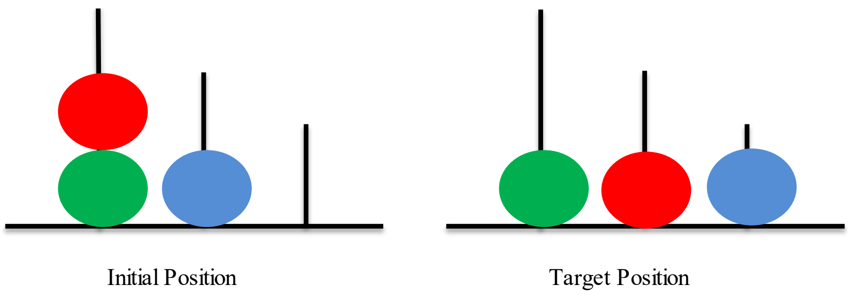

Figure 1: A subproblem of the Tower of London Test. The initial position is the same for all problems.

Ward & Allport (1997) have claimed that the original version of TOL may be useful with special populations, but it is too simple for investigating normal subjects' planning ability. The variations of TOL have been developed and differed according to problem sets, number, and colors of discs, task instructions (Tunstall, 1999). There are 3-, 4- and 5-disc versions of TOL (Shallice,1982; Kafer & Hunter, 1997; Ward & Allport, 1997). Besides time limits and rule breaks, the variables that were altered affect test scores in different versions of TOL. Tunstall (1999) mentioned several limits of TOL with different variations. These are;

- The time limit applied in some versions causes confusion about measuring the planning ability or planning speed.
- The unstated rationale for problem selection.
- The ability to discriminate colors affects test performance.

Tunstall (1999) aimed to eliminate the limitations listed above by developing the 4-disc version of TOL. Tunstall (1999) increased the number of discs from 3 to 4 to overcome a commonly observed ceiling effect in the original version (Shum, Ungvari, Tang & Leung, 2004). In addition, Tunstall (1999) altered discs' colors, enabling individuals with color vision deficiencies to discriminate the discs (see figure 2).

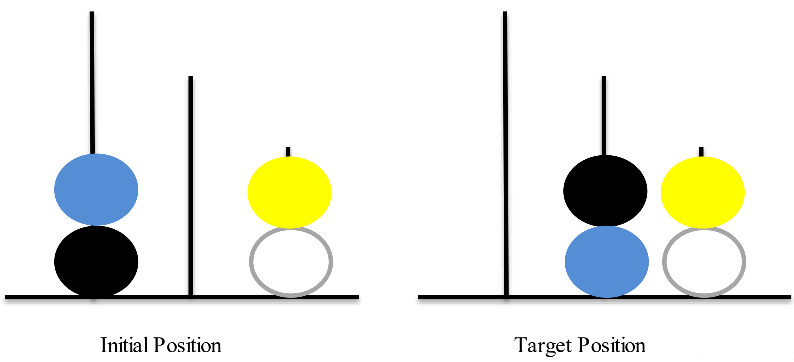

Figure 2: A subproblem of the 4 disc TOL. The initial position is the same for all problems.

The features of 4-disc TOL, which eliminated the ability of color discriminating, increased level of test difficulty, its improved scoring method, and its promising psychometric properties are the reasons to be preferred by this study and others (Shum et al., 2009; Andrews et al., 2014).

Atalay & Cinan (2007) conducted a reliability and validity study of the Drexel version of the Tower of London test (three pegs, red, green, and blue) for adult groups in Turkish society. However, the features of the 4-disc TOL mentioned above provide advantages for use in diverse groups. The 4-disc TOL was used by Cinan & Ünsal (2011) to investigate the impact of the performance of the perceptual properties of the beads on planning performance in Turkish society. In this study, the computerized version of the 4-disc TOL was developed and applied in Turkish society.

Computers provide the advantage of more precise measurement, timings, and presentation speeds that can be controlled precisely, especially when complex cognitive skills are being assessed (Singleton, Horne & Simmons, 2009). Computerized measurements might represent potential cost savings, not only with regards to materials and supplies but also the time that the test administrator needs (Wild, Howieson, Webbe, Seelye & Kaye, 2008). Assessing older children and adults can be generally self-administered, and results can be obtained immediately; both these factors help reduce administrative workload and avoid retardation (Singleton, Horne & Simmons, 2009). Reduced need for administration by trained personnel provides cost savings and scheduling flexibility as benefits of computerized testing (Wild et al., 2008). Besides, the subjective judgment of the test administrator does not affect the test results and reliability of measurement (Singleton, Horne & Simmons, 2009).

Singleton, Horne & Simmons (2009) claimed that adults, as well as children, often prefer computerized assessment to traditional assessment, and participants often find traditional assessment by another person stressful and anxiety-provoking, particularly when the assessor is perceived as 'teacher' or some equivalent professional. Results of research carried out by Fillit, Simon, Doniger & Cummings (2008) about the practicality of the computerized system for cognitive assessment showed that patients have found the computerized assessment easy to use and understand. Computerized measures to investigate planning ability certainly have practical advantages, such as the precise timing of stimulus presentation and automatic recording of behavioral responses (Tecwyn, Thorpe, & Chappell, 2013).

Despite the advantages mentioned above, there are a limited number of computerized assessments and planning ability measurements in Turkish literature. Cinan (2015) developed Istanbul 5 Cube Planning Tower measuring the planning ability and computerized version. However, long-standing literature accumulation and clinical studies provide advantages in measuring planning skills with TOL.

In this study, a computerized version was developed and gathered data on child samples differently from Cinan & Ünsal (2011). At this point, the current study provides advantages to be used in further studies and extensive samples. The study aimed to determine item discrimination, difficulty, reliability levels of a computerized version of 4-disc TOL and measure planning skills in different age groups and gender.

METHOD

Research Design

The current study was designed as a quantitative general screening model.

Research Sample

The current study was comprised of three studies that have different participants described below in the procedure section. TOL was transformed into a computerized test form, but it was not applied via an online platform; instead, it was applied offline in the environment created by researchers in study 1. Since the research was conducted in a wide age range, more participants were needed in each age group. For this reason, the computerized test has been moved to the online platform. Thus, study 2 with a children sample and study 3 with an adult sample were conducted online. Due to the elapsed time and different test presentations, the data collected later were not combined with the first collected data but were planned as separate studies. Table 1 is presented to express the sample's distribution in different phases of the study more clearly.

Table 1. Study sample

Phases	Age		Gender	
	5-14	15-53	Female	Male
Study 1 (combined sample)	82	170	129	123
Study 2 (children sample)	88	-	41	47
Study 3 (adult sample)	-	244	134	110
Total	170	414	304	280

The total number of participants in the study was 584. Participants aged 5-14 years attend primary school (N:127) and secondary school (N:43) in the 2016-2017 academic year in Samsun. The education level of participants aged 15-53 years were high school (N:100), graduate school (N:292), and postgraduate (N:22). The detailed demographic data about the participants were reported in each study below.

Research Instruments and Procedures

4-disc version of the Tower of London was developed by Tunstall (1999). Besides, reliability and validity studies of TOL for clinical use were conducted by Tunstall, O'gorman & Shum (2016). Developing a new version of TOL was conducted with three experiment processes by Tunstall (1999). Tunstall initially created a large pool of possible items that varied in the number of moves required to solve, and the most sensitive items were selected from the pool regarding item difficulty and discriminating psychometric criteria (Shum et al., 2009). The internal consistency level of 4-disc TOL was compared with 3-disc TOL and was found better. The test-retest reliability level of the test was found acceptable. Besides, provided normative data towards ages 5 – 53 within a heterogeneous group, reliability and validity study results showed that 4-disc TOL measures planning ability properly. Tunstall used Krikorian, Bartok & Gay's (1994) scoring method on 4-disc TOL, which involves removing possible speed and accuracy confusions and giving the maximum potential for discrimination by including three trials per item (Shum et al., 2009). 4-disc TOL consists of 10 items that are split in half as "simple" problems require 2 to 5 moves, and "complex" problems require 6 to 9 moves (Shum et al., 2009). TOL includes a total score, the number of problems solved on the first trial, the number of problems solved within three trials, and planning time (Tunstall, 1999).

The total score consists of total points in which three points were awarded for solving a problem on the first trial, two points on the second trial, one point on the third trial, and zero if the problem was not solved within three trials (Tunstall, 1999). The total score can be obtained from the 4-disc TOL ranging from 0 to 30. The score of the number of problems solved on the first trial measures the accuracy of initial plans, which means accuracy runs fairly straightforward through the planning process (Tunstall, 1999). The score of the number of problems solved within three trials measures the ability to monitor actions, use feedback to modify incorrect plans, and adjust plans (Tunstall, 1999). Planning time was calculated as the average time from the commencement of the trial until the first disc was released and summed across the first trials of all tried problems (Tunstall, 1999). Planning time score measures patterning abilities such as conceptualization, generation of visuospatial patterns, possible solutions, analysis, and inhibiting inappropriate responses (Tunstall, 1999).

The computerized version of 4-disc TOL was developed by using the Adobe Flash Professional CS6 program. The introduction of the test was made via a video record showing test rules, sample problems, and solutions. Data about participants' responses, such as the number of solved problems, movements, and movement time, were saved via the used online platform automatically. The problems in the original version developed by Tunstall (1999) were used in the computerized version, and any new planning problem was not constructed in this study. The validity and reliability study of the 4-disc TOL was conducted by Tunstall (1999). In this study, there was no need for a validity study other than item analysis since no new items were added and there were no cultural elements in the items. Item discrimination, difficulty, and internal consistency levels of 4-disc TOL in this research are presented in the Results section.

Research data had been collected in two phases. In the first phase, the computerized version of the 4-disc TOL was developed and applied to participants aged 5-53 years (N:252). We communicated with children sample through primary schools and applied tests in state primary and secondary schools. The data towards the adult sample also were gathered in state schools and a university in Samsun. Six months later, in the second phase, the computerized version of 4-disc was published on the website at oyna.biliminrenklidunyasi.com to collect more data. This website was not open to access, and it was just used to collect data. At this phase, participants were asked to use the test via the website. In the second phase, 88 children (aged 5-14 years) and 244 adults (15-53) participated. Participants were asked to indicate where they live in the online form attached to the TOL. Thus, participants who live in Samsun were included in the study. Research data were evaluated as in three studies because of the varying data collection method in two phases.

Data Analysis

The normality of study data was tested primarily. Kolmogorov Smirnov normality test was used to determine if study data sets were well-modeled by a normal distribution. Test results showed that TOL scores were not distributed normally within age and gender groups ($p<.05$). In addition, transformations did not make the distribution acceptably normal. Thus, the Kruskal- Wallis H test as a nonparametric test was used to compare TOL scores according to age groups. In order to assess the effects of age on planning skills properly, research samples were divided into age groups. Age groups were designated based upon reliability and validity studies about TOL (Boccia et al., 2017; Tunstall, 1999; Tunstall et al., 2016) and to ensure that the number of participants in the groups was suitable for group comparison. Age groups were 5-9, 10-14, 15-17, 18-25, 26-35, 36-53 years in combined and adult samples. However, in order to assess in detail age group was designated in a smaller range 5-6, 7-8, and 9-14 years in the children sample similar to the validity study of TOL by Tunstall (1999).

Mann Whitney U with Bonferroni Correction was used to find out the group, which caused significant differences between age groups. The critical p-value was divided by the number of comparisons made (α/k) for Bonferroni correction. The statistical results of the tests were then calculated based on these modified p values. Mann-Whitney U test was used to compare TOL scores according to the gender variable. The effect size was calculated for statistically significant test results. Cohen (1988) has provided benchmarks to define small ($\eta^2 = 0.01$; d=.2), medium ($\eta^2 = 0.06$; d=.5) and large ($\eta^2 = 0.14$; d=.8) effects. Item difficulty and item discrimination indexes were compared with the Z test, and Cohen's h was calculated for statistically significant results.

RESULTS

Study 1

In this study, it was aimed to compare the TOL measures according to age and gender variables in a sample comprised of children and adults. The combined sample comprised 252 participants (Female:129, Male:123) aged between 5-9 years, 10-14 years, 15-17 years, 18-25 years, 26-35 years, 36-53 years.

Table 2. Descriptive statistics towards TOL scores according to age

		Total Score		Planning Time		FT		TT	
Age	N	x̄	SD	x̄	SD	x̄	SD	x̄	SD
5-9 years	47	16.7	4.2	1.7	1.2	4.4	1.4	2.1	1.3
10-14 years	35	17.8	4.3	1.7	0.8	4.6	1.6	2.2	1.1
15-17 years	41	20.8	2.9	1.7	0.9	5.8	1.2	2.1	1.2
18-25 years	39	20	4.7	1.9	1.9	5.6	1.2	1.9	0.9
26-35 years	44	22.5	2.5	2.5	3.4	6.6	1.3	1.6	1.3
36-53 years	46	20.6	4	2	1.9	6	1.6	1.5	1.1

FT: Number of problems solved on the first trial

TT: Number of problems solved within three trials

Table 3. Correlation between TOL scores and age

	N	Total Score			Planning Time			FT			TT		
		r	p	r^2	r	p	r^2	r	p	r^2	r	p	r^2
Age	252	.357**	.000	.12	.002	.979	-	.389**	.000	.15	-.194**	.002	.03

**p<.01

Spearman's rank-order correlation test showed that there was a statistically significant relationship between the total score and age variable (r=.357, p<.01), which indicates a moderate positive correlation (Cohen, 1988). The age variable explained 12% of the variation in the total score. There was no statistically significant relationship between planning time and age variable (r=.002, p>.01). There was a statistically significant relationship between FT and age variable (r=.389, p<.01), which indicates a moderate positive correlation with the age variable explaining 15% of the variation in FT score. There was a statistically significant relationship between TT and the age variable (r=-.194, p<.01), which indicates a small negative correlation (Cohen,1988). The age variable explained 3% of the variation in TT score.

Table 4. Comparison of TOL scores according to the age

	Age	N	Mean Rank	df	χ^2	p	Effect Size η^2	d_{cohen}	Bonferroni	p	Effect Size η^2	d_{cohen}
Total Scoree	5-9 years (1)	47	73.4	5	56.15	.000*	.20	1	1-3	.000**	.26	1.18
	10-14 years (2)	35	92.6						1-4	.001**	.13	.78
	15-17 years (3)	41	142.1						1-5	.000**	.46	1.8
	18-25 years (4)	39	135.4						1-6	.000**	.19	.96
	26-35 years (5)	44	175.2						2-3	.002**	.12	.76
	36-53 years (6)	46	138.2						2-5	.000**	.34	1.4
Planning Time	5-9 years (1)	47	124.2	5	1.543	.908						
	10-14 years (2)	35	137.7									
	15-17 years (3)	41	125.9									
	18-25 years (4)	39	118									
	26-35 years (5)	44	130.1									
	36-53 years (6)	46	124.4									
FT	5-9 years (1)	47	76.9	5	57.6	.000*	.21	1	1-3	.000**	.21	1
	10-14 years (2)	35	87.3						1-4	.001**	.13	.78
	15-17 years (3)	41	136.9						1-5	.000**	.40	1.6
	18-25 years (4)	39	132.4						1-6	.000**	.22	1
	26-35 years (5)	44	174.9						2-3	.001**	.14	.81
	36-53 years (6)	46	146.2						2-5	.000**	.31	1.3
									2-6	.000**	.15	.86
TT	5-9 years (1)	47	138.6	5	12.96	.024*	.03	.36				
	10-14 years (2)	35	144.1						1-6	.033		
	15-17 years (3)	41	137.7						2-6	.009		
	18-25 years (4)	39	132.6						3-6	.022		

26-35 years (5)	44	107.8		1-5	.054
36-53 years (6)	46	103.2		2-5	.024

*p<.05,**p<0.008

Kruskal-Wallis H test showed that there was a statistically significant difference in the total score between the age groups and the age variable has a large effect size on total score variance $\chi^2(5) = 56.15$, p = .000, η^2=.20, Cohen's d= 1. Mann Whitney U with Bonferroni Correction was made to find out the group that caused significant differences between age groups. Mann Whitney U with Bonferroni Correction test result showed that there were significant differences in total score between age groups that 5-9 and 15-17 years U=390.5, p=.000; 5-9 and 18-25 years U=525, p=.001; 5-9 and 26-35 years U=211, p=.000; 5-9 and 36-53 years U= 534, p=.000; 10-14 and 15-17 years U= 417.5, p;=.002, 10-14 and 26-35 years U=242.5, p=.000. Kruskal-Wallis H test showed that there was no statistically significant difference in planning time between the age groups $\chi^2(5) = 1.543$, p = .908.

Kruskal-Wallis H test showed that there was a statistically significant difference in FT between the age groups and age variable has large effect size on FT variance χ2(5) = 57.6, p = .000, η^2=.21, Cohen's d= 1. Mann Whitney U with Bonferroni Correction test result showed that there were significant differences in FT between age groups that 5-9 and 15-17 years U= 446.5, p=.000, 5-9 and 18-25 years U= 525.5, p= .001; 5-9 and 26-35 years U= 274, p=.000; 5-9 and 36-53 years U= 482.5, p=.000; 10-14 and 15-17 years U= 403.5, p=.001; 10-14 and 26-35 years U= 263.5, p=.000; 10-14 and 36-53 years U= 429, p=.000.

Kruskal-Wallis H test showed that there was a statistically significant difference in TT between the age groups, and the age variable has a small effect size on TT variance χ2(5) = 12.96, p = .024, η2=.03, Cohen's d=0.36. Although the Kruskal- Wallis H test result, there was no significant difference in TT between age groups when compared with Mann-Whitney U with Bonferroni Correction test, p>.008.

Table 5. Comparison of TOL scores according to gender

	Gender	N	Mean Rank	Sum of Rank	U	z	p	Effect Size	
								η^2	d_{cohen}
Total Score	Female	129	123.3	15915.5	7330.5	-.699			
	Male	123	129.3	15962.5			.484	-	
Planning	Female	129	116.2	14995	6610	-2.288			
Time	Male	123	137.2	16883			.022*	.02	.29
FT	Female	129	125.7	16223.5	7838.5	-.167			
	Male	123	127.2	15654.5			.867	-	
TT	Female	129	124.7	16012.5	7627.5	-.546			
	Male	123	128.9	15865.5			.585	-	

p<.05

Mann-Whitney U test result showed that there was no significant difference in total score, FT, and TT between female and male participants (U=7330, p>.05). Besides, it was found that there was a significant difference in planning time between female and male participants (U=6610, p<.05), and gender had a small effect size on planning time variance. Moreover, male participants had a higher planning time mean rank than female participants.

Study 2

In this study, it was aimed to compare the TOL measures according to age and gender variables in the children sample. The sample comprised 88 children (girl:41, boy:47) aged between 5-14 years (8.6±2.5).

Table 6. Descriptive statistics towards TOL scores according to age

		Total Score		Planning Time		FT		TT	
Age	N	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
5-6 years	20	16.2	4.3	1.9	1	4.4	1.4	1.7	1
7-8 years	32	17.1	4.4	1.8	1.5	4.4	1.5	2.2	1.4
9-14 years	36	18	4.4	1.8	0.8	4.7	1.7	2.1	1.1
Total	88	17.3	4.4	1.8	1.1	4.5	1.5	2.1	1.2

Table 7. Comparison of TOL scores according to age

	Age	N	Mean Rank	df	χ^2	p
Total Score	5-6 ages	20	38.6	2	2.795	.247
	7-8 ages	32	42.3			
	9-14 ages	36	49.6			
Planning Time	5-6 ages	20	47.8	2	.985	.611
	7-8 ages	32	41.1			
	9-14 ages	36	45.6			
FT	5-6 ages	20	42.2	2	.566	.753

	7-8 ages	32	43.2			
	9-14 ages	36	46.8			
TT	5-6 ages	20	38.1	2	1.783	.410
	7-8 ages	32	47.3			
	9-14 ages	36	45.5			

p<.05

A Kruskal-Wallis H tests results showed that there was no statistically significant difference between the age groups according to total score $\chi^2(2)$= 2.795, p=.247, planning time $\chi^2(2)$ = .985, p = .611; number of problems solved on the first trial $\chi^2(2)$ = .566, p = .753; number of problems solved within three trials $\chi^2(2)$ = 1.783, p = .410.

Table 8. Descriptive statistics towards TOL scores according to gender

			Total Score		Planning Time		FT		TT	
	Gender	N	x̄	SD	x̄	SD	x̄	SD	x̄	SD
Children Sample	Girl	41	18	3.9	1.7	0.9	4.5	1.5	2	1.1
	Boy	47	16.6	4.7	1.8	1.3	4.3	1.6	2.1	1.3
Total		88								

Table 9. Comparison of TOL scores according to gender

		Gender	N	Mean Rank	Sum of Rank	U	z	p
Total Score	5-6 years	Girl	11	11.18	123	42	-.573	.566
		Boy	9	9.67	87			
	7-8 years	Girl	7	19.57	137	66	-.988	.323
		Boy	25	15.64	391			
	9-14 years	Girl	23	18.98	436.5	138.5	-.365	.715
		Boy	13	17.65	229.5			
Planning Time	5-6 years	Girl	11	10.23	112.5	46.5	-.228	.820
		Boy	9	10.83	97.5			
	7-8 years	Girl	7	17.21	120.5	82.5	-.228	.820
		Boy	25	16.30	407.5			
	9-14 years	Girl	23	17.26	397	121	-.939	.347
		Boy	13	20.69	269			
FT	5-6 years	Girl	11	11.32	124.5	40.5	-.697	.486
		Boy	9	9.50	85.5			
	7-8 years	Girl	7	19.36	135.5	67.5	-.937	.349
		Boy	25	15.70	392.5			
	9-14 years	Girl	23	18.96	436	139	-.351	.726
		Boy	13	17.69	230			
TT	5-6 years	Girl	11	10.14	111.5	45.5	-.318	.750
		Boy	9	10.94	98.5			
	7-8 years	Girl	7	17.50	122.5	80.5	-.327	.744
		Boy	25	16.22	405.5			
	9-14 years	Girl	23	17.89	411.5	135.5	-.478	.633
		Boy	13	19.58	254.5			

p<.05

Mann-Whitney U test results showed that there was no statistically significant difference between girls and boys according to total score, planning time, the number of problems solved on the first trial, and the number of problems solved within three trials.

Study 3

In this study, it was aimed to compare the TOL measures according to age and gender variables in the adults sample. The sample comprised 244 adults (Female:134, Male:110) aged between 15-53 years (28.1±12.3).

Table 10. Descriptive statistics towards TOL scores according to age

		Total Score		Planning Time		FT		TT	
Age	N	x̄	SD	x̄	SD	x̄	SD	x̄	SD
15-17 years	59	19.6	3.4	1.6	0.8	5.3	1.4	2.2	1.2
18-25 years	76	20.8	4.1	1.7	1.5	5.9	1.6	1.8	1.1
26-35 years	57	22.1	2.5	2.3	3	6.5	1.3	1.6	1.3
36-53 years	52	20.4	4	2.1	1.9	5.9	1.6	1.6	1.1

Table 11. Comparison of TOL scores according to age

	Age	N	Mean Rank	df	χ^2	p	Effect Size η^2	Effect Size d_{cohen}	Bonferroni	p	Effect Size η^2	Effect Size d_{cohen}
Total Score	15-17 years (1)	59	98.7	3	14.46	.002*	.04	.44	1-3		0.12	0.76
	18-25 years (2)	76	128							.002**		
	26-35 years (3)	57	146.5									
	36-53 years (4)	52	114.9									
Planning Time	15-17 years (1)	59	121.3	3	1.198	.756						
	18-25 years (2)	76	117.4									
	26-35 years (3)	57	122.4									
	36-53 years (4)	52	131.1									
FT	15-17 years (1)	59	96.7	3	16.852	.001*						
	18-25 years (2)	76	123.6				.05	.49				
	26-35 years (3)	57	149.3						1-3	.000**	0.14	0.83
	36-53 years (4)	52	120.6									
TT	15-17 years (1)	59	141.1	3	9.602	.022*	.02	.33				
	18-25 years (2)	76	127.8						1-3	.009**	0.05	0.48
	26-35 years (3)	57	107									
	36-53 years (4)	52	110.1									

*p<.05, **p<0.0125

Kruskal-Wallis H test showed that there was a statistically significant difference in a total score between the age groups χ2(3) = 14.460, p = .002, η2=.04, Cohen' d=.44, and the age variable have a small effect size on total score variance. Mann Whitney U with Bonferroni Correction test result showed a significant difference in a total score between 15-17 years and 26-35 years age groups U=982, p=.002. Results showed that the total score mean rank of adults who are 26-35 years old was higher than the total score mean rank of adults who are 15-17 years old.

Kruskal-Wallis H test showed no statistically significant difference in planning time between the age groups χ2(3) = 1.198, p = .756.

Kruskal-Wallis H test showed a statistically significant difference in the number of problems solved on the first trial between the age groups χ2(3) = 16.852, p = .001, η2=.05, Cohen' d=.49, and the age variable have a small effect size on FT variance. Mann Whitney U with Bonferroni Correction test results showed a significant difference in the number of problems solved on the first trial between 15-17 years and 26-35 years U=931.5, p=.000. Results showed that the FT score mean rank of adults who are 26-35 years old was higher than the FT score mean rank of adults 15-17 years old.

Kruskal-Wallis H test showed a statistically significant difference in the number of problems solved within three trials (TT) between the age groups χ2(3) = 9.602, p = .022, η2=.02, Cohen' d=.33, and the age variable have a small effect size on TT variance. Mann Whitney U with Bonferroni Correction test results showed that there was a significant difference in TT score between 15-17 years and 26-35 years U=1223, p=.009. Results showed that the TT score mean rank of adults who are 15-17 years old was higher than the TT score mean rank of adults 26-35 years old.

Table 12. Descriptive statistics towards TOL scores according to gender

			Total Score		Planning Time		FT		TT	
Age Groups	Gender	N	x̄	SD	x̄	SD	x̄	SD	x̄	SD
15-17 years	Female	38	19.5	3.6	1.5	0.7	5.3	1.6	2.1	1.2
	Male	21	19.8	3	1.7	1	5.3	1.2	2.3	1.2
18-25 years	Female	40	20.5	4	1.5	1.6	5.7	1.5	1.9	1.1
	Male	36	21.2	4.2	1.9	1.5	6	1.7	1.7	1.1
26-35 years	Female	29	22.1	2.4	2	3	6.5	1.1	1.6	1.5
	Male	28	22.2	2.7	2.5	3.1	6.5	1.5	1.5	1.2
36 -53 years	Female	27	19.4	3.3	1.8	2.1	5.5	1.3	1.7	1.1
	Male	25	21.4	4.6	2.3	1.6	6.4	1.8	1.4	1.1

Table 13. Comparison of TOL scores according to gender

		Gender	N	Mean Rank	Sum of Rank	U	z	p	Effect Size η^2	Effect Size d_{cohen}
Total Score	15-17 years	Female	38	29.37	1116	375	-.382	.703		
		Male	21	31.14	654					
	18-25 years	Female	40	35.25	1410	590	-1.359	.174		
		Male	36	42.11	1516					
	26-35 years	Female	29	28.29	820.5	385.5	-.330	.741		
		Male	28	29.73	832.5					
	36-53 years	Female	27	22.09	596.5	218.5	-2.189	.029*	.09	.63

		Male	25	31.26	781.5					
Planning Time	15-17 years	Female	38	29.41	1117.50	376.5	-.356	.722		
		Male	21	31.07	652.50					
	18-25 years	Female	40	33.28	1331	511	-2.175	.030*	.06	.51
		Male	36	44.31	1595					
	26-35 years	Female	29	25.52	740	305	-1.612	.107		
		Male	28	32.61	913					
	36-53 years	Female	27	23.43	632.5	254.5	-1.521	.128		
		Male	25	29.82	745.5					
FT	15-17 years	Female	38	29.96	1138.5	397.5	-.024	.981		
		Male	21	30.07	631.5					
	18-25 years	Female	40	35.78	1431	611	-1.159	.246		
		Male	36	41.53	1495					
	26-35 years	Female	29	28.43	824.5	389.5	-.273	.785		
		Male	28	29.59	828.5					
	36-53 years	Female	27	22.46	606.5	228.5	-2.037	.042*	.07	.57
		Male	25	30.86	771.5					
TT	15-17 years	Female	38	28.86	1096.5	355.5	-.710	.478		
		Male	21	32.07	673.5					
	18-25 years	Female	40	39.78	1591	669	-.553	.580		
		Male	36	37.08	1335					
	26-35 years	Female	29	29.36	851.5	395.5	-.173	.862		
		Male	28	28.63	801.5					
	36-53 years	Female	27	29.04	784	269	-1.306	.191		
		Male	25	23.76	594					

p<.05

Mann-Whitney U test results showed that there was a significant difference between the total score of participants who are 36-53 years old according to gender variable U=218.5, p=.029, η2=.09, Cohen's d=.63. Total mean rank score of male participants is higher than female participants in the 36-53 years age group, and the gender variable has a medium effect size on the total score in that age group. Mann-Whitney U test results showed a significant difference between the planning time of participants who are 18-25 years old according to gender variable U=511, p=.030, η2=.06, Cohen's d=.51. Mean rank planning time of male participants is higher than female participants in the 18-25 years age group, and the gender variable has a medium effect size on planning time in that age group. Mann-Whitney U test results showed that there was a significant difference in FT of participants who are 36-53 years old according to gender variable U=228.5, p=.042, η2=.07, Cohen's d=.57. Mean rank FT of male participants are higher than female participants in the 36-53 years age group, and the gender variable has a medium effect size on FT in that age group. Mann-Whitney U test results showed that there was no significant difference between TT of participants according to gender variable.

Table 14. Comparison of item difficulty and item discrimination levels of children and adult data

Item No	p_j^1	p_j^2	Z	Cohen h	r_{jx}^1	r_{jx}^2	Z	Cohen h
1	0.94	0.95	-0.382	-	0.12	0.07	1.52	-
2	0.73	0.86	-2.88*	0.30	0.44	0.24	3.72*	0.44
3	0.84	0.93	-2.57*	0.25	0.20	0.13	1.66	-
4	0.84	0.93	-2.57*	0.25	0.27	0.15	2.62*	0.28
5	0.65	0.84	-3.92*	0.44	0.34	0.33	0.18	-
6	0.70	0.84	-2.97*	0.34	0.55	0.29	4.61*	0.56
7	0.61	0.83	-4.40*	0.50	0.65	0.34	5.36*	0.62
8	0.32	0.55	-3.95*	0.48	0.56	0.73	-3.11*	0.36
9	0.07	0.10	-0.90	-	0.08	0.21	-2.99*	0.36
10	0.08	0.14	-1.58	-	0.13	0.32	-3.74*	0.46
Average	0.58	0.70	-2.17*	0.22	0.33	0.28	0.94	-

p_j^1, r_{jx}^1: Children Data, p_j^2, r_{jx}^2: Adult Data, p<.05

According to Table 14, items of TOL represent a range of difficulty from .08 to .94 for the children sample. According to Güler (2017), the item difficulty index ranges from 0 to 1, and .50 is mid-range. Besides, it is concerned that items difficulty index less than .20: too difficult, .40 - .60: excellent, and more than .90: too easy (Quaigrain & Arhin, 2017). Accordingly, TOL has an average medium difficulty level in children's data. Besides, items of TOL represent a range of difficulty from .10 to .95 and have an easy item difficulty level for the adult sample. Z test results towards a comparison of difficulty indexes showed that the average item difficulty level for the adult sample was significantly higher than the children sample average item difficulty. It means that TOL was easier for the adult sample than the children sample, and the sample variable has a small effect size on TOL item difficulty. It is possible to see the same results within other items.

Table 14 shows that items of TOL represent a range of discrimination from .08 to .65 for the children sample. It is concerned that item discrimination .40 and more: very good, .30 - .39: good, .20-.29: fairly good, .19 and less: poor (Ebel; 1979, cited by Quaigrain & Arhin, 2017). Accordingly, TOL's average item discrimination level is good for the children sample. Besides, TOL represents a range of discrimination from .07 to .73 for the adult sample, and the average item discrimination index level is fairly good. Z test results towards a comparison of item discrimination indexes showed that there was no significant difference between the two sample's item discrimination levels. The items with 2nd, 4th, sixth, and seventh numbered have higher item discrimination levels in the children sample than the adult sample. However, the 8th, 9th, and 10th numbered items have higher item discrimination levels in the adult sample than the children sample.

Table 15. Internal consistency reliability of TOL

N	Cronbach Alpha	Split Half
252	.60	.62

p<.05

Table 15 shows that the internal reliability Cronbach Alpha coefficient of TOL was .60, and split-half reliability was found .62. Flynn, Schroeder & Sakakibara (1994) stated that a Cronbach's alpha of 0.60 and above was considered an acceptable reliability level (as cited in Ringim, Razalli, & Hasnan; 2012). Besides, Anastasi & Urbina (1997) suggested using split-half reliability in case of increasing difficulty between items.

DISCUSSION, CONCLUSION, AND RECOMMENDATIONS

Results of the first study conducted with a combined sample showed that there were significant relationships between TOL scores (total score, FT, and TT) and the age variable. However, there was no significant relationship between planning time and age. Besides, there were significant differences between age groups regarding total score, FT, and TT. No significant difference was found between the total score, FT, and TT of female and male participants. However, there was a significant difference in planning time regarding the gender variable in the combined sample.

Results of the second study conducted with children sample showed that there was no significant difference between age groups regarding each TOL score. Besides, there was a significant difference between TOL scores according to the gender variable.

Results of the third study conducted with adult samples showed that there were significant differences between age groups regarding total score, FT, and TT. No significant difference was found between age groups by planning time. Besides, there were significant differences between the total score, planning time, and FT of female and male participants. There was no significant difference between TT scores of adults by gender.

Research results were interpreted by combining the results of three studies. Firstly, TOL score results according to the age and gender variables were interpreted. Secondly, item analysis and internal consistency results were mentioned.

Research results showed that there was a significant positive relationship between total score and age. It is possible to interpret this result as increasing age variable was associated with an increasing total score or decreasing age was associated with decreasing the total score. In other words, the planning ability changes depending on age. In literature, it was stated that the planning develops until the individual reaches adulthood (Becker et al., 1987; Dreher & Oerter, 1987; Welsh et al., 1991 as cited by Mahapatra, 2016). Oku & Aihara (2008) stated that the prefrontal cortex is involved in many complex cognitive functions such as problem-solving, planning, reasoning, and decision-making. The age-related developments in executive functions have been associated with the maturation of the prefrontal cortex (Diamond, 2002). Besides, planning ability may be relevant to the ability to keep information in mind. Planning ability in TOL requires keeping in mind the following peg to be moved. Case et al. (1982) found that people could keep in mind more words by age (as cited by Diamond, 2002). The metanalysis study of frontal functioning showed that age is an important predictor of variations in frontal functions such as planning ability (Romine & Reynolds, 2005). The current study finding is compatible with previous researches investigated planning ability according to the age variable. Krikorian et al. (1999) found that total score improved in a linear progression from younger to older age groups. Besides, previous research result showed statistically significant linear and quadratic trends in TOL scores with increasing age (Albert & Steinberg, 2011). Similarly, Tunstall (1999) had found that the 4-disc TOL total score increased by age. Phillips, Kliegel, & Martin (2006) had reached a similar result with the study finding on the age-related increase in planning ability. There are studies reporting that performance on the TOL improved with age (Albert & Steinberg, 2011; Asato, Sweeney, & Luna, 2006; Malloy-Diniz et al., 2008).

Research results showed a statistically significant difference between total scores according to the age groups, and the age variable has a large effect size on planning ability variance. There were significant differences in planning ability levels of participants between 5-9 and 15-17 years; 5-9 and 18-25 years; 5-9 and 26-35 years; 5-9 and 36-53 years; 10-14 and 15-17 years; and 10-14 and 26-35 years. Mean rank planning ability scores of children who were 5-9 years old were significantly lower than older ages, excluding 10-14 ages. Besides, planning ability scores in 10-14 ages were significantly lower than in older ages. Similarly, research results showed that there was a statistically significant positive relationship between the number of problems solved on the first trial score and the age variable and a significant difference between the age groups. It was found that the age variable has a large effect size on FT variance. There were significant differences in the number of problems solved on the first trial score between age groups that 5-9 and 15-17 years; 5-9 and 18-25 years; 5-9 and 26-35 years; 5-9 and 36-53 years; 10-14 and 15-17 years; 10-14 and 26-35 years; and 10-14 and 36-53 years. Mean rank FT scores of children who were 5-9 years old were significantly lower than older ages, excluding 10-14 ages. Besides, FT scores in 10-14 ages were significantly lower than in older ages. Also, when compared with age-related changes in the children sample, there was no significant difference in planning ability and FT between small age ranges.

These results may be based on the ongoing improvement process of planning ability until the last stages of childhood. The prefrontal cortex, associated with cognitive skills like planning ability, is the last brain region to mature with full maturation, which is not reached until 16 years of age (Fuster,1999; Tunstall, 1999). These findings show a similarity with the findings of the study of Luciana, Collins, Olson, & Schissel (2009). When compared with age-related changes in planning ability in the adult sample, results showed that the total score mean rank of adults in the 26-35 years old age group was higher than adults in the 15-17 years old age group. This is consistent with a rapidly advancing capacity for planning during the childhood years that reaches a peak in early adulthood (Tunstall, O'gorman, & Shum, 2016). Research results are compatible with the previous result that TOL performance improves with age but levels out in the adult years reported by Albert and Steinberg (2011).

Research results showed that there was no significant relationship between planning time and age. Besides, there was no statistically significant difference in planning time between the age groups. Also, no significant difference was found in planning time according to age compared to children and adult samples separately. Although they were not significant statistically, compared mean rank results showed that the middle age group took shorter planning time than younger and older age groups. Tunstall (1999) indicated that slowing in planning time occurs at around 20 years. The younger and older age groups took equal planning time. Similarly, Tunstall (1999) had reported that planning time decreased from childhood to adolescence but increased for adults. Besides, planning time leveled off at 10-14 years. Similarly, Luciana et al. (2009) reported that TOL planning times leveled off at age 15.

Research results showed a significant negative relationship between the number of problems solved within three trials score and age. Besides, there was a statistically significant difference in the number of problems solved within three trials scores between the age groups, and the age variable has a small effect size on variance. However, there was no significant difference when compared with age groups in the combined sample. This result may be based on the fact that Narum (2006) stated that Bonferroni correction may be too conservative and results in greatly diminished power to detect differentiation among pairs of sample collections. No significant difference was found in the children sample by age, too. However, when compared separately, there were statistically significant differences between the number of problems solved within three trials scores of adults according to the age group variable.

The number of problems solved within three trials score of adults in 15-17 years age group was higher than 26-35 years age group. This finding showed consistency with the result of a negative relationship between age and the number of problems solved within three trials score. The youngest age group in the adult sample had a higher number of problems solved within three trials score than older groups. Results showed that the number of problems solved within three trials scores increased with decreasing age. In other words, younger adults had more tendency to use feedback, monitoring actions than older adults. Older adults are often worse than younger adults at adapting to changing situational demands, and this difference is commonly attributed to an age-related decline in acquiring and updating information (Wilson, Nusbaum, Whitney, & Hinson, 2017).

Research results showed no significant difference in planning ability, in the number of problems solved on the first trial score, in the number of problems solved within three trials score between male and female participants. Researchers reported no significant difference in TOL scores between males and females (Krikorian et al., 1994; Atalay & Cinan, 2007; Masson, Dagnan, & Evans, 2010). However, when compared with age groups in the adult sample, there was a significant difference between the total score and the number of problems solved on the first trial scores of participants who are 36-53 years old according to gender variable. Total mean rank score and the number of problems solved on the first trial scores of male participants were higher than female participants in 36-53 years age group. Boccia et al. (2017) indicated that there were gender differences in performing TOL. While males showed higher precuneus activity, suggesting that they relied on visuospatial abilities, females showed higher activity in the dorsolateral prefrontal cortex, suggesting that they relied more on executive processing (Boghi et al., 2006). Asato et al. (2006) reported that there were gender differences in TOL4 performance, and males performed better than females in the adult group.

In the present study, there was no significant difference in planning ability, planning time, the number of problems solved on the first trial scores, the number of problems solved within three trials scores between boys and girls in the children sample. Asato et al. (2006) claimed that this gender effect was not present at younger ages indicating that this bias may be established after adolescence. There are researches reporting that there was no significant difference between performances of the children by gender (Culbertson & Zillmer; 1998; Asato et al., 2006).

In the present study, there was a significant difference in planning time between male and female participants. Male participants took more time than female participants while planning. When the source of significant difference is searched, it was found that there was a significant difference between the planning time of participants who are 18-25 years old according to the gender variable in the adult sample. Mean rank planning time of male participants was higher than female participants in the 18-25 years age group. Boccia et al. (2017) reported that there was a significant difference in planning time by gender, suggesting that by the age of 26, there is an overturn in the gender differences in the speed of planning. There was no significant difference between the TT of participants according to the gender variable.

Item analysis results showed that difficulty and discrimination levels of items increase based upon increasing moves. This result is similar to the findings of the study conducted by Tunstall (1999). Kaller, Unterrainer & Stahl (2012) reported a clear and nearly perfect linear increase in task difficulty by moves. Item reliability analysis showed that Cronbach's Alpha and split-half reliability coefficients are higher than the initial study of developing 4-disc TOL by Tunstall (1999).

Research results showed that planning ability is correlated with increasing age. In further studies, planning skills in daily life activities that stand out in individuals in diverse age groups who have different levels of planning skills can be investigated. Research results showed that male participants' planning ability and the number of problems solved on the first trial scores were higher than female participants in 36-53 age groups. In connection with this result, gender roles that may affect the planning ability with increasing age may be investigated in next studies. In further studies, TOL might be used within diverse and extended samples. In this study, only item difficulty, discrimination, and internal consistency analysis were performed, and studies including detailed validity and reliability analysis can be conducted in future studies.

Limitations

The ability and predisposition of elderly participants to point with mouse and touchpad should be considered as a limitation as a criterion not excluded in the study.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Statements of publication ethics

We hereby declare that the study has not unethical issues and that research and publication ethics have been observed carefully. E.G.D conceived of the presented idea and developed the theory, and performed the computations. Y.Ö. encouraged E.G.D to investigate the research topic and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Researchers’ contribution rate

The study was conducted and reported with equal collaboration of the researchers.

Ethics Committee Approval Information

Ethics committee approval was obtained from Ondokuz Mayıs University Social and Human Sciences Ethics Committee with the protocol number 2016/97 and dated 15/07/2016.

REFERENCES

Albert, D., & Steinberg, L. (2011). Age differences in strategic planning as indexed by the Tower of London. Child Development, 82(5), 1501–1517.

Allain, P., Nicoleau, S., Pinon, K., Etcharry-Bouyx, F., Barré, J., Berrut, G., … Le Gall, D. (2005). Executive functioning in normal aging: A study of action planning using the Zoo Map Test. Brain and cognition, 57(1), 4–7.

Anatasi, A. & Urbina, S. (1997).Psychological testing (7th ed.) Upper Saddle River, NJ: Prentice Hall.

Andrews, G., Halford, G. S., Chappell, M., Maujean, A., & Shum, D. H. (2014). Planning following stroke: A relational complexity approach using the Tower of London. Frontiers in human neuroscience, 8, 1032.

Asato, M. R., Sweeney, J. A., & Luna, B. (2006). Cognitive processes in the development of TOL performance. Neuropsychologia, 44(12), 2259–2269.

Atalay, D., & Cinan, S. (2007). Yetiskinlerde planlama becerisi: Londra Kulesi (LK^ sup DX^) testinin standardizasyon ve güvenilirlik çalışması. Türk Psikoloji Dergisi, 22(60), 25.

Boccia, M., Marin, D., D’Antuono, G., Ciurli, P., Incoccia, C., Antonucci, G., … Piccardi, L. (2017). The Tower of London (ToL) in Italy: standardization of the ToL test in an Italian population. Neurological Sciences, 38(7), 1263–1270.

Boghi, A., Rasetti, R., Avidano, F., Manzone, C., Orsi, L., D’agata, F., … Pulvirenti, L. (2006). The effect of gender on planning: An fMRI study using the Tower of London task. Neuroimage, 33(3), 999–1010.

Cinan, S., & Ünsal, P. (2011). Renk-şekil boyutu değiştirme ve ayırt etme süreçlerinin planlamaya etkisi. Türk Psikoloji Dergisi, 26(68).

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (Second Edition). Lawrence Erlbaum Associates.

Culbertson, W. C., & Zillmer, E. A. (1998). The Tower of LondonDX: A standardized approach to assessing executive functioning in children. Archives of Clinical Neuropsychology, 13(3), 285–301.

Das, J. P., & Georgiou, G. K. (2016). Levels of planning predict different reading comprehension outcomes. Learning and Individual Differences, 48, 24–28.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. Principles of frontal lobe function, 466–503.

Dockery, C. A., Hueckel-Weng, R., Birbaumer, N., & Plewnia, C. (2009). Enhancement of planning ability by transcranial direct current stimulation. Journal of Neuroscience, 29(22), 7271–7277.

Fillit, H. M., Simon, E. S., Doniger, G. M., & Cummings, J. L. (2008). Practicality of a computerized system for cognitive assessment in the elderly. Alzheimer's & Dementia, 4(1), 14-21.

Fossati, P., Ergis, A. M., & Allilaire, J. F. (2002). Executive functioning in unipolar depression: a review. L'Encephale, 28(2), 97–107.

Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in "planning" functions? Interpreting data from the Tower of Hanoi. Neuropsychologia, 33(5), 623–642.

Güler, N. (2017). Eğitimde Ölçme ve Değerlendirme (10.Baskı). Ankara: Pegem Akademi.

Harlow, J. M. (1869). Recovery from the passage of an iron bar through the head. D. Clapp.

Hughes, C. (2002). Executive functions and development: Emerging themes. Infant and Child Development, 11(2), 201–209.

Kafer, K. L., & Hunter, M. (1997). On testing the face validity of planning/problem-solving tasks in a normal population. Journal of the International Neuropsychological Society, 3(2), 108–119.

Kaller, C. P., Unterrainer, J. M., & Stahl, C. (2012). Assessing planning ability with the Tower of London task: Psychometric properties of a structurally balanced problem set. Psychological Assessment, 24(1), 46.

Kopecky, H., Chang, H. T., Klorman, R., Thatcher, J. E., & Borgstedt, A. D. (2005). Performance and private speech of children with attention-deficit/hyperactivity disorder while taking the Tower of Hanoi test: Effects of depth of search, diagnostic subtype, and methylphenidate. Journal of Abnormal Child Psychology, 33(5), 625–638.

Krikorian, R., Bartok, J., & Gay, N. (1994). Tower of London procedure: a standard method and developmental data. Journal of clinical and Experimental Neuropsychology, 16(6), 840–850.

Less, A. (2008). Relations between preschool children's planning ability, self-regulation and early literacy skills.

Luciana, M., Collins, P. F., Olson, E. A., & Schissel, A. M. (2009). Tower of London performance in healthy adolescents: The development of planning skills and associations with self-reported inattention and impulsivity. Developmental Neuropsychology, 34(4), 461–475.

Mahapatra, S. (2016). Development of Planning Behaviour and Decision Making Ability of Children. Journal of Education and Practice, 7(6), 74–77.

Malloy-Diniz, L. F., Cardoso-Martins, C., Nassif, E. P., Levy, A. M., Leite, W. B., & Fuentes, D. (2008). Planning abilities of children aged 4 years and 9 months to 8 1/2 years: Effects of age, fluid intelligence and school type on performance in the Tower of London test. Dementia & Neuropsychologia, 2(1), 26–30.

Masson, J. D., Dagnan, D., & Evans, J. (2010). Adaptation and validation of the Tower of London test of planning and problem solving in people with intellectual disabilities. Journal of Intellectual Disability Research, 54(5), 457–467.

McCormack, T., & Atance, C. M. (2011). Planning in young children: A review and synthesis. Developmental Review, 31(1), 1–31.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. Annual review of neuroscience, 24(1), 167–202.

Narum, S. R. (2006). Beyond Bonferroni: less conservative analyses for conservation genetics. Conservation genetics, 7(5), 783–787.

Oku, M., & Aihara, K. (2008). A mathematical model of planning in the prefrontal cortex. Artificial Life and Robotics, 12(1–2), 227–231.

Owen, A. M. (1997). Cognitive planning in humans: neuropsychological, neuroanatomical and neuropharmacological perspectives. Progress in neurobiology, 53(4), 431–450.

Pea, R. D. (1982). What is planning development the development of? New Directions for Child and Adolescent Development, 1982(18), 5–27.

Phillips, L. H., Kliegel, M., & Martin, M. (2006). Age and planning tasks: The influence of ecological validity. The International Journal of Aging and Human Development, 62(2), 175–184.

Rönnlund, M., Lövdén, M., & Nilsson, L. G. (2001). Adult age differences in Tower of Hanoi performance: Influence from demographic and cognitive variables. Aging, Neuropsychology, and Cognition, 8(4), 269-283.

Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. Cogent Education, 4(1), 1301013.

Ringim, K. J., Razalli, M. R., & Hasnan, N. (2012). A framework of business process re-engineering factors and organizational performance of Nigerian banks. Asian Social Science, 8(4), 203.

Romine, C. B., & Reynolds, C. R. (2005). A model of the development of frontal lobe functioning: Findings from a meta-analysis. Applied neuropsychology, 12(4), 190–201.

Rönmark, L. (2014). The Never Ending Shower: planning ability, intellectual disability and cognitive artifacts.

Shallice, T. (1982). Specific impairments of planning. Phil. Trans. R. Soc. Lond. B, 298(1089), 199–209.

Shum, D., Gill, H., Banks, M., Maujean, A., Griffin, J., & Ward, H. (2009). Planning ability following moderate to severe traumatic brain injury: Performance on a 4-disk version of the Tower of London. Brain Impairment, 10(3), 320–324.

Shum, D., Ungvari, G. S., Tang, W.-K., & Leung, J. P. (2004). Performance of schizophrenia patients on time-, event-, and activity-based prospective memory tasks. Schizophrenia Bulletin, 30(4), 693.

Simon, H. A. (1975). The functional equivalence of problem solving skills. Cognitive Psychology, 7(2), 268–288.

Singleton, C., Horne, J., & Simmons, F. (2009). Computerised screening for dyslexia in adults. Journal of Research in Reading, 32(1), 137-152.

Tecwyn, E. C., Thorpe, S. K., & Chappell, J. (2013). A novel test of planning ability: Great apes can plan step-by-step but not in advance of action. Behavioural processes, 100, 174–184.

Tunstall, J. R. (1999). Improving the utility of the Tower of London: A neuropsychological test of planning. Griffith University Brisbane,, Australia.

Tunstall, J. R., O'gorman, J. G., & Shum, D. H. (2016). A four-disc version of the Tower of London for clinical use. Journal of neuropsychology, 10(1), 116–129.

Unterrainer, J. M., Rahm, B., Halsband, U., & Kaller, C. P. (2005). What is in a name: Comparing the Tower of London with one of its variants. Cognitive Brain Research, 23(2–3), 418–428.

Ward, G., & Allport, A. (1997). Planning and problem solving using the five disc Tower of London task. The Quarterly Journal of Experimental Psychology Section A, 50(1), 49–78.

Wild, K., Howieson, D., Webbe, F., Seelye, A., & Kaye, J. (2008). Status of computerized cognitive testing in aging: a systematic review. Alzheimer's & Dementia, 4(6), 428-437.

Wilson, C. G., Nusbaum, A. T., Whitney, P., & Hinson, J. M. (2017). Age-differences in cognitive flexibility when overcoming a preexisting bias through feedback. Journal of clinical and experimental neuropsychology, 1–9.