Published in final edited form as:
Dev Neuropsychol. 2009 ; 34(4): 461–475. doi:10.1080/87565640902964540.

Tower of London Performance in Healthy Adolescents: The Development of Planning Skills and Associations With Self-Reported Inattention and Impulsivity

Monica Luciana, **Paul F. Collins**, **Elizabeth A. Olson**, and **Ann M. Schissel**
Department of Psychology and Center for Neurobehavioral Development University of Minnesota, Minneapolis, Minnesota

Abstract

Studies have investigated planning skill development using the Tower of London (TOL). Reports conflict regarding maturational trajectories and associations with IQ, other executive functions, and impulsivity. A convenience sample of 9- to 20-year-olds completed the TOL and other measures. TOL accuracy improved until ages 15–17. Digit span backwards (DSB), response inhibition, and IQ were correlated with TOL performance. DSB contributed to TOL accuracy above and beyond age and IQ. Inhibitory control and DSB both contributed to the modulation of planning times across problems. Self-reported inattention and hyperactivity were associated with low performance. Task approaches reflecting planning and psychometric issues are discussed.

Conventional wisdom suggests that effective planning is crucial to self-organization. It involves setting a goal, formulating a checklist of tasks necessary to achieve it, and executing each one until the goal is achieved. Anecdotal reports suggest that children and adolescents are perceived as deficient in planning skills, which is not surprising given that executive functions improve markedly through adolescence (Luciana, Conklin, Hooper, & Yarger, 2005; Luna, Garver, Urban, Lazar, & Sweeney, 2004). The planning involved in real-life contexts is difficult to measure experimentally (Phillips, Kliegel, & Martin, 2006), but one frequently used planning task is the Tower of London (TOL: Shallice, 1982). It requires individuals to manipulate objects to conform to visible spatial solutions. Task variants can be distinguished by their physical structure, the problem space sampled, and scoring techniques. Most incorporate measures of planning accuracy and planning time, which is the time taken to initiate a problem. Non-computerized versions quantify rule-breaking errors (Berg & Byrd, 2002). Because each trial's solution is visible, one critique of the task is that the goal does not have to be internalized to succeed on a given problem. An individual can work step-by-step using trial and error feedback to assess each move's consequences (Purcell, Maruff, Kyrios, & Pantelis, 1998), implying that the TOL recruits perceptual and visuospatial, as well as higher cognitive, functions. Nonetheless, performance is compromised with frontal lobe impairment, in impulsive behavior disorders,

Correspondence should be addressed to Monica Luciana, Department of Psychology, 75 East River Road, University of Minnesota, Minneapolis, MN 55455. lucia003@umn.edu.

and in severe psychopathology (Badcock, Michiel, & Rock, 2005; Owen, Downes, Sahakian, Polkey, & Robbins, 1990; Rhodes, Coghill & Matthews, 2005).

Given the importance of planning for self-organized behavior, the TOL has been used in developmental studies. Performance improves throughout childhood and into adolescence (Krikorian, Bartok, & Gay, 1994; Luciana & Nelson, 1998, 2002). Luciana and Nelson (2002) used the Cambridge Neuropsychological Testing Automated Battery (CANTAB) to measure executive functions in healthy 4- to 12-year-olds as compared to college students, and reported that 12-year-olds had not reached young adult levels of TOL performance in terms of perfect solutions and number of moves to complete five-move problems. De Luca et al. (2003) studied 8- to 64-year-olds using the CANTAB and reported that 8- to 14-year-olds achieved fewer perfect solutions than 15- to 29-year-olds. The groups were maximally different in solving three- and five-move problems. Planning times were not reported in these studies. Asato, Sweeney, and Luna (2006) reported negative associations between age and the number of moves to complete four- and five-move CANTAB TOL problems in healthy 8- to 30-year-olds as well as age-related increases in initiation times. Age also impacted the execution times for three- and four-move problems. The sample was aggregated into child (ages 8–13), adolescent (ages 14–17), and adult (ages 18–30) groups. All age-related differences were due to differences between children and adolescents and/or adults with no differences between the latter groups. However, adults scored one standard deviation below children and adolescents in full-scale IQ scores, which may have attenuated performance differences between adolescents and adults, because planning skills have been predicted by fluid intelligence (Unterrainer et al., 2004; Zook, Welsh, & Ewing, 2006).

Asato et al. (2006) also found that better TOL performance was associated with the ability to suppress inappropriate eye movements during an antisaccade task. Smaller correlations were reported between the TOL and an oculomotor delayed response measure of working memory (WM). These findings support that the TOL is associated with inhibitory control (IC) (Welsh, Satterlee-Cartmell, & Stine, 1999) as well as WM (Gilhooly, Wynn, Phillips, Logie, & Della Sala, 2002; Welsh et al., 1999; Zook, Davalos, DeLosh, & Davis, 2004). The TOL/oculomotor task correlations were not moderated by age.

Whether age-related increases in planning are due to similar increases in other executive functions has not been well-investigated. Huizinga, Dolan, and van der Molen (2006) examined the contributions of WM, shifting, and IC to TOL performance using structural equation modeling. Each cognitive domain was represented by three tasks. TOL planning times leveled off at age 15. The number of perfect TOL solutions increased until age 21. However, there was not a reliable multivariate model that could predict TOL summary variables using latent variables generated from other behavioral domains. For the prediction of the number of perfect solutions, Stroop task performance was correlated with TOL accuracy only in 21-year-olds, suggesting that IC impacts performance only in adulthood. But for 7- to 15-year-olds, the 95% confidence intervals for the regression coefficients could not be calculated, and latent measures of shifting and WM cohered in factor analyses, but measures of inhibition did not. Thus the role of IC in the development of planning remains unclear. Other disparities exist as well. Some studies have reported better TOL performance in adult males (Asato et al., 2006); others have found no sex effects (DeLuca et al., 2003; Krikorian et al., 1994; Luciana & Nelson, 2002).

Thus, controversy remains regarding when planning performance asymptotes, whether performance differences are evident between adolescence and young adulthood, whether the nature of age trends varies for accuracy versus planning times, how the task's timing variables relate to performance, and how strongly the task is impacted by other executive functions. A less extensively investigated issue concerns whether TOL performance maps onto other indicators of planning or impulsivity. The current study addresses each of these areas.

METHODS

Participants

Two convenience samples (combined N = 323, ages 9 to 20 years) were tested across consecutive years. The first study (N = 152; 79 male, 73 female) employed behavioral measures. The second study (N = 171; 93 male, 78 female) used the same measures as well as structural neuroimaging (not reported here). In both studies, participants under the age of 18 were recruited from a database maintained by the University of Minnesota's Institute for Child Development. When their children were born, parents were identified through birth records and invited to join the database. We invited consenting families with children between the ages of 9 and 17 years were to participate in a study of adolescent brain development. Inclusion criteria included being a native English speaker, having normal/ corrected-to-normal vision and hearing, and having no history of neurological or psychological illness, mental retardation, or learning difficulties. These criteria were assessed by telephone and verified by in-person interviews of the participant and parent (see later). All 18- to 20-year-olds were recruited through on-campus advertisements. Selection criteria were the same as described earlier. In the first study, the criteria were assessed through a telephone screening of the parent (or adult participant) followed by a direct interview using an in-house questionnaire. In the second study, this questionnaire was followed with parent and child structured interviews (Kaufmann et al., 1997). Intelligence was measured in the first study by the Vocabulary and Block Design subtests of the Wechsler Intelligence Scale for Children–III (ages 9 to 16: Wechsler, 1991) or the Wechsler Adult Intelligence Scale–III (ages 17 or higher: Wechsler, 1997a). Intelligence was measured in the second study by the four subtests of the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999). Participants and parents provided informed consent/ assent according to local IRB requirements. The samples were equivalent in all measures reported herein, so they were combined (Table 1).

Procedures

Testing occurred on a separate day. The TOL was one of several cognitive tasks, including Go-No-Go and digit span tasks, that were administered. A subset of study two's participants (N = 116, aged 11 to 20 years) completed self-report versions of the Conners Rating Scale–Revised (CRS:R:L: Conners, 2000), which yields an attention deficit hyperactivity disorder (ADHD) index score and inattention and hyperactivity/impulsivity subscale scores.

Tower of London—This computerized task is referred to as *the Stockings of Cambridge* within the Cambridge Neuropsychological Testing Automated Battery (CANTAB: Sahakian & Owen, 1992). It was administered using a personal computer and touch-screen monitor. A full task description can be found in Luciana and Nelson (1998). The task consists of a problem-solving block and a yoked following block. Using a touch-screen, participants move colored balls to match a target display (problem-solving block) or to mimic moves initiated by the computer (yoked following block). Participants were told at the start of each problem-solving trial that the trial should be completed in *X* number of moves, where *X* is the minimum number of moves required to achieve a perfect solution. The starting position of the balls varied. Across trials, the solution could be reached after two, three, four, or five moves. The time before making the first move and the subsequent execution time were recorded.

On each trial of the yoked condition, the computer executes moves, and the participant imitates them. This task is yoked to the problem-solving portion in that for each trial of the yoked condition, the balls'movements are exact replicas of moves made by the participant in the corresponding problem-solving trials (Veale, Sahakian, Owen, & Marks, 1996). These movement times allow planning and execution times to be derived for the main task. For each problem, the planning time is the time between the presentation of the problem and the first touch minus the corresponding initiation time from the yoked task. When planning or initiation times are referenced throughout this article, we are referring to this difference score. The execution time is also a difference score, calculated as the time between selecting the first move and problem completion minus the corresponding times from the yoked task. The task consists of 2 two-move problems, 2 three-move problems, 4 four-move problems, and 4 five-move problems. The total number of perfect solutions was recorded. For each problem set, the average number of moves to complete the set, planning times, and execution times were calculated.

Go-No-Go task—The Go-No-Go task (Braver, Barch, Gray, Molfese, & Snyder, 2001) was programmed in E-Prime (Psychology Software Tools, www.psnet.com). Participants viewed letters (20% Xs) presented one-by-one on a computer screen. Stimulus duration was 250 msec with a 1,000 msec inter-stimulus interval. In the first no-go trial block, participants responded to all letters except X and inhibited responses to Xs. A second target detection block demanded attentional vigilance but not motor inhibition. Instructions were to press a button only in response to Xs (20% occurrences).

Digit memory span—The digit span task (Wechsler, 1997b) was administered to yield measures of forward and backward verbal memory span. The span was the largest number of items that could be successfully remembered in the correct (forward or backward) order.

RESULTS

Data were analyzed using the Statistical Package for the Social Sciences, version 11.5. Age effects on TOL performance were assessed using analyses of variance. Measures of IQ, WM, and response inhibition were correlated with performance. Hierarchical regressions assessed whether associated processes contributed to task performance above and beyond the other measures. Associations between levels of inattentive/impulsive behavior (CRS) and TOL performance were also determined. Alpha levels of .05 were used to determine statistical significance. Effect sizes are represented by η_p^2, r, and R^2 values.

Age-Related Trends in Performance (see Table 2)

Age groups were created to fully represent adolescence (9–11-year-olds: $N = 92$; 12–14-year-olds: $N = 82$; 15–17-year-olds: $N = 81$; 18–20-year-olds: $N = 65$). TOL variables were examined using mixed model analyses of variance, with difficulty (performance on two-, three-, four-, and five-move problems) as a within subjects variable and age group as a between-subjects variable. Difficulty by age group interactions were followed up with one-way analyses of variance. Other summary variables were compared between age groups by one-way analyses of variance. The Tukey HSD procedure was used to determine the nature of observed age effects.

Accuracy—Analysis of the average number of moves to complete each problem set revealed a main effect of difficulty level, $F(3, 948) = 2016.29$, $p < .001$, $\eta_p^2 = .87$, a main effect of age group, $F(3, 316) = 18.97$, $p < .001$, $\eta_p^2 = .15$, and a difficulty by age group interaction, $F(9, 948) = 7.13$, $p < .001$ $\eta_p^2 = .06$. Age group effects were found for all sets (η_p^2 values were .03, .04, .09, and .10 for two-, three-, four-, and five-move problems). For two-move problems, 15–17-year-olds performed worse than 12–14- ($p = .03$) and 18–20-year-olds ($p = .04$), but as Table 2 indicates, most individuals performed perfectly. For three-move problems, 9–11-year-olds performed worse than 18–20-year-olds ($p = .02$).. For four-move problems, 9–11-year-olds performed worse than all other groups (all ps ≤ .02), which were equivalent. For five-move problems, 9–11-year-olds were worse than 15–17- and 18–20-year-olds (ps ≤ .001). The 12–14- year-olds were worse than 18–20-year-olds ($p = .004$). The older two groups were equivalent. The number of perfect task solutions varied by age, $F(3, 316) = 21.07$, $p < .001$, $\eta_p^2 = .17$. The 9–11-year-olds were worse than all other groups (ps < .001). The 12–14-year-olds were worse than 18–20-year-olds ($p = .002$). The 15–17-year-olds were not different from 18–20-year-olds ($p = .08$). Participant sex did not impact these findings.

Planning times—Analysis of planning times across problem sets revealed a main effect of difficulty, $F(3, 948) = 191.07$, $p < .001$, $\eta_p^2 = .38$, no main effect of age group, and a difficulty by age group interaction, $F(9, 948) = 7.12$, $p < .001$, $\eta_p^2 = .06$. The age groups varied in planning times for three- and five-move problems (η_p^2 values were .05 and .06). For three-move problems, 9–11-year-olds were slower than 12–14- and 15–17-year-olds (ps = .05 and .00, respectively), but not 18–20-year-olds ($p = .06$). For five-move problems, 9–11-year-olds were *faster* than 15–17- and 18–20-year-olds (ps = .008 and .001). Sex did not moderate these findings.

Given the pattern of planning times for five-move problems, we assessed whether longer planning times supported better performance on the TOL. Planning times were positively correlated with the number of perfect solutions, $r(318) = .32$, $p < .001$. To examine how this association varied by age, an index (Number of Perfect Solutions × Average Planning Time) was computed, reflecting the extent to which planning time contributed to maximal performance. This score increased with age, $F(3, 316) = 3.69$, $p = .01$, $\eta_p^2 = .03$. The 18–20-year-olds had higher scores than 9–11-year-olds ($p = .01$).

Because planning times were computed as difference scores (see Methods), increases could reflect increments in deliberation prior to starting a problem or decrements in the time to initiate movements in the yoked condition. Analyses of yoked initiation times yielded a significant effect of difficulty, $F(3, 948) = 31.72$, $p < .001$, $\eta_p^2 = .09$, a significant effect of age group, $F(3, 316) = 25.66$, $p < .001$, $\eta^{p2} = .20$, but no significant interaction, $F(15, 942) = 0.79$, $p = .49$. The main effect of difficultywas due to slower movement initiation times for two-move problems than for all other sets (all *p*s < .001). Movement times for three-move sets were slower than for four- ($p = .000$) and five-move sets ($p = .045$). The age group effect was due to 9–11- year-olds being slower than all other groups (*p*s < .001).

When the initiation times (i.e., uncorrected by subtraction of the corresponding times from the yoked condition) were examined, there was a main effect of difficulty level, $F(3, 948) = 179.32$, $p < .001$, $\eta_p^2 = .36$, no effect of age group, $F(3, 316) = 0.79$, $p = .50$, and an interaction between age group and difficulty level, $F(9, 948) = 6.79$, $p < .001$, $\eta_p^2 = .06$. The effect of difficulty was due to shorter initiation times for two-move relative to three-, four-, and five-move problems (all *p*s < .001). Three-move problems elicited shorter initiation times than did four- and five-move problems, and four-move problems elicited shorter initiation times than five-move problems (all *p*s < .001). The age by difficulty interaction was due to age effects on initiation times to two-, three-, and five-move problems (η_p^2 values = .03, .07, and .04). For two-move problems, 9–11-year-olds had longer latencies to begin the problems than did 15–17-year-olds, $p = .02$. For three-move problems, 9–11-year-olds had longer latencies to initiate than all older age groups, all *p*s ≤ .02. For five-move problems, 9–11-year-olds had shorter initiation times than 15–17- ($p = .03$) and 18–20-year-olds ($p = .003$). Thus, the age-by-difficulty interaction for planning times appears to reflect cognitive versus motor processes.

Execution times—Analysis of execution times (see Methods) yielded a main effect of difficulty, $F(3, 948) = 120.93$, $p < .001$, $\eta_p^2 = .28$, a main effect of age, $F(3, 316) = 4.65$, $p = .003$, $\eta_p^2 = .04$, and an age by difficulty interaction, $F(9, 948) = 3.12$, $p = .001$, $\eta_p^2 = .03$. *T*-tests indicated that two-move problems were solved faster than all other problem sets, three-move problems were solved faster than four- and five-move problems, and four-move problems were solved *slower* than five-move problems (all *p*s < .001). The Age × Difficulty interaction was due to group differences in executing three- and four-move problems (η_p^2 = .04 and .06, respectively). For three move problems, 12–14-year-olds were faster than 15–17- and 18–20- year-olds (*p*s = .003 and .03). For four-move problems, 9–11-year-olds took more time than 15–17- and 18–20-year-olds (*p*s = .000 and .02). No other differences were significant.

Proportion of problem solution time due to planning—We next considered whether age impacts the relative amount of time allotted to planning versus execution across sets. For each problem, a ratio was calculated, dividing uncorrected initiation time by total problem time (initiation plus execution time). These scores (see Table 2) indicate that more than half of a problem's overall solution time reflected planning, and upon analysis, they showed an effect of difficulty, $F(3, 945) = 79.32$, $p < .001$, $\eta_p^2 = .20$, an effect of age group, $F(3, 315) = 8.97$, $p < .001$, $\eta_p^2 = .08$, and a difficulty by age group interaction, $F(9, 945) = 11.45$, $p < .001$, $\eta_p^2 = .10$. For two-move problems, there was less time spent planning relative to three-, four-, and five-move problems (all *p*s < .001); for three-move problems, there was relatively more time spent planning as compared to four-move ($p < .001$) but not five-move problems; and for four-move problems, there was less time spent planning than for five-move problems ($p < .001$). The age effect was due to 9–11- and 12–14-year-olds spending less overall solution time planning versus executing as compared to 15–17- and 18–20-year-olds (*p*s < .02). In terms of the interaction, there were no age effects for two- or three-move problems. Age impacted four-move problems, $F(3, 315) = 8.00$, $p < .001$, $\eta_p^2 = .07$, with 9–11- and 12–14-year-olds both spending relatively less time planning relative to 15–17- and 18–20-year-olds. Age also impacted five-move problems, $F(3, 315) = 24.64$, $p < .001$, $\eta_p^2 = .19$. The 9–11-year-olds spent less relative time planning than did all other groups; the 12–14- year-olds spent less time planning than 15–17- and 18–20-year-olds. The older groups did not differ from one another. We considered the slope of this performance feature, that is, the extent to which the proportion of time devoted to planning versus execution increased linearly from two-to-three-to-four-to-five-move problems for each participant. The slope showed a strong age effect, $F(3, 315) = 25.11$, $p < .001$, $\eta_p^2 = .19$. The 9–11-year-olds had lower values than all other groups; 12–14-year-olds exhibited lower values than 15–20-year-olds.

Correlations Between Timing Variables and Accuracy

The average execution time across difficulty levels was positively correlated with the average planning time, $r(318) = .25$, $p < .001$. Individuals who took more time to start a problem took more time to complete it. This correlation was carried most by easier problem sets and was not significant for the most difficult problems: two-move, $r(318) = .21$, $p < .001$; three-move, $r(318) = .21$, $p < .001$; four-move: $r(318) = .12$, $p = .04$; and five-move problems: $r(318) = .10$, $p = .07$. In addition, increased planning times benefited performance as indicated by a positive correlation between the average planning time and perfect task solutions, while increased execution times reflected poorer performance, $r(318) = -.29$, $p < .001$. When execution times were correlated with the average number of moves to complete each problem set, longer times were associated with greater numbers of moves (two-move: $r(318) = .32$, $p < .001$; three-move: $r(318) = .42$, $p < .001$; four-move: $r(318) = .49$, $p < .001$; five-move: $r(318) = .34$, $p < .001$). This pattern held within each age group. In contrast, increased planning times within sets correlated with more accurate performance on four-, $r(318) = -.13$, $p = .02$, and five-move, $r(318) = -.32$, $p < .001$, problems.

The planning time slope correlated with perfect solutions, $r(318) = .52$, $p < .001$, with the mean number of moves to complete four- ($r(318) = -.30$, $p < .001$) and five-move ($r(318) = -.53$, $p < .001$) problems, and with the accuracy-by-planning time index, $r(318) = .47$, $p < .001$.

Associations with IQ

IQs ranged from 80 to 150 and were weakly associated with perfect solutions, $r(314) = .18$, $p = .001$, with three-move planning times, $r(314) = -.13$, $p = .03$, execution times for three-, four-, and five-move problems (each $r(314) = -.14$, $ps \leq .02$), and moves to complete three-, $r(314) = -.18$, $p = .002$, and four-move problems, $r(314) = -.11$, $p = .04$.

Associations with Memory Span

Digit span was evaluated in a mixed model ANOVA with two levels of process (forward, backward). There was an effect of age group, $F(3, 317) = 37.75$, $p < .001$, $\eta_p^2 = .26$, an effect of process, $F(1, 317) = 426.7$, $p < .001$, $\eta_p^2 = .57$, but no interaction between the two. Forward span exceeded backward span. The 9–11-year-olds had lower spans than all other groups ($ps < .001$). The 12–14-year-olds had lower spans than 15–17- ($p = .004$) and 18–20-year-olds ($p < .001$). The 15–17-year-olds had lower spans than 18–20-year-olds ($p = .05$). Sex was unrelated to digit span.

Positive associations were found between forward digit span and the number of perfect TOL solutions, $r(316) = .25$, $p < .001$, and average moves to complete three- ($r(316) = -.17$, $p = .003$), four- ($r(316) = -.21$, $p < .001$), and five-move ($r(316) = -.14$, $p = .01$) problems. It was also correlated with planning times for three- ($r(316) = -.11$, $p = .049$) and five-move ($r(316) = .11$, $p = .046$) problems as well as with the planning time proportion slope variable ($r(316) = .21$, $p < .001$). Backward span was correlated with number of perfect solutions, $r(316) = .39$, $p < .001$, average moves to complete three-, $r(316) = -.17$, $p = .003$, four, $r(316) = -.25$, $p < .001$, and five-move problems, $r(316) = -.30$, $p < .001$, and planning times for three-, $r(316) = -.12$, $p = .03$, four-, $r(316) = .12$, $p = .04$, and five-move problems, $r(316) = .23$, $p < .001$, average planning times, $r(316) = .14$, $p = .02$, the Planning Time × Accuracy index score, $r(316) = .23$, $p < .001$, average execution times, $r(316) = -.15$, $p < .01$, and the slope variable, $r(316) = .30$, $p < .001$.

Associations With Measures of Attention and Response Inhibition (Go-No-Go Task)

Hit and false alarm rate variables were log-transformed prior to analysis. For vigilant attention, age effects were found on target detection reaction times (RTs), $F(3, 312) = 62.20$, $p < .001$, $\eta_p^2 = .37$, hit rates, $F(3, 312) = 10.83$, $p < .001$, $\eta_p^2 = .09$, and false alarm rates, $F(3, 312) = 29.84$, $p < .001$, $\eta_p^2 = .22$. RTs decreased until ages 15–17. Hit and false alarm rates both improved until ages 12–14. In terms of response inhibition, RTs to “go” stimuli during the no-go block were impacted by age, $F(3, 312) = 46.42$, $p < .001$, $\eta_p^2 = .31$, as were false alarm rates, $F(3, 312) = 64.20$, $p < .001$, $\eta_p^2 = .38$. RTs and false alarm rates both decreased until ages 15–17.

Associations between the TOL and Go-No-Go variables were examined excluding one outlier who responded opposite to the task rules. Target detection and no-go false alarm rates were related to poorer TOL performance as measured by perfect task solutions, $rs(311) = -.27$ and $-.30$, respectively; both $ps < .001$). High false alarm rates on both conditions were also associated with more moves to complete three-, four-, and five-move problems (rs ranged from .14 to .27, all $ps \leq .01$, $dfs = 311$). High target detection and no-go false alarm rates were also associated with shorter planning times for five-move problems, $rs(311) = -.15$ and $-.19$, $ps < .01$, and with lower values of the proportion slope variable, $rs(311) = -.30$ and $-.39$, $ps < .001$.

High target detection hit rates were associated with greater numbers of perfect solutions, $r(311) = .21$, $p < .001$. For moves to complete three, four- and five-move problems, a higher target detection hit rate also related positively to performance, *r*s(311) = −.13, −.15, and −.15 respectively; *p*s =.03, .01, and .01. High hit rates were also associated with high values of the proportion slope variable, $r(311) = .24$, $p < .001$. Faster target detection RTs were associated with higher numbers of perfect solutions, $r(311) = -.28$, $p < .001$, as were Go-No-Go RTs, $r(311) = -.19$, $p = .001$. Slower go trial and target detection RTs were associated with greater numbers of moves to complete three-, four-, and five-move problems (*r*s ranged from .13 to .23).

Multivariate Prediction of TOL Performance

We next predicted TOL performance on the basis of age, IQ, WM, attention and IC using hierarchical regression analyses, incorporating TOL dependent variables that reflected global aspects of task performance: the number of perfect TOL solutions, the average planning time, the Accuracy × Planning Time index, and the planning time proportion slope variable. For each dependent variable, a three-step regression analysis was performed, first entering age and IQ, then Go-No-Go and target detection hit and false alarm rates, and then the two digit span variables.

The regression model to predict average planning time was not significant. However, in the prediction of the planning time proportion slope variable, the first step (age and IQ) was significant with an R^2 of .19. Both increased age and higher IQs were significant predictors of a high slope value. The second step yielded a significant increment in R^2, $F(6, 306) = 15.26$, $p < .001$, $R^2 = .23$. Significant predictors were age ($t = 4.04$, $\beta = .27$), IQ ($t = 1.98$, $\beta = .10$), and no-go false alarm rate ($t = –2.55$, $\beta = -.17$). The third step (digit span variables) yielded a non-significant ($p = .057$) increase in R^2, $F(8, 306) = 12.31$, $p < .001$, $R^2 = .25$. Within this model, age and no-go false alarm rate remained as significant predictors along with digit span backward ($t = 2.34$, $\beta = .14$), but not full-scale IQ. In the prediction of perfect solutions the first step was significant with an R^2 of .22, $F(2, 306) = 42.48$, $p < .001$. Age ($t = 8.59$, $\beta = .44$) and IQ ($t = 4.32$, $\beta = .22$) were significant predictors. The increment in R^2 on the second step was not significant. The third step was significant with an overall R^2 of .27, $F(8, 306) = 13.65$, $p < .001$. Significant predictors were age ($t = 4.15$, $\beta = .30$), IQ ($t = 2.94$, $\beta = .16$), and backward digit span ($t = 3.97$, $\beta = .23$). For the Planning Time × Accuracy index, only digit span backward ($t = 3.01$, $\beta = .20$) remained as a significant predictor after the third step, $F(8, 306) = 3.24$, $p = .001$, $R^2 = .08$.

Interactive age effects were explored in additional regressions on the number of perfect solutions and the Planning Time × Accuracy score. The first step entered age and IQ, the second step entered backward digit span, and the third step entered the interactive term of age multiplied by backward digit span. The interaction term was not a significant predictor of either variable. A similar analysis was conducted for the planning time slope proportion variable that also included age by no-go false alarm rate interactions. Again, the interaction term was not significant.

Associations With Measures of Inattention and Hyperactivity

Raw CRS scores were converted to age and gender-scaled T-scores. The T-score range for the inattention index was 31 to 71 (M= 42.4, SD = 8.7); the range for the hyperactivity index was 28 to 74 (M = 39.5, SD = 9.8). The number of perfect TOL solutions correlated negatively with self-reported inattention, $r(114) = -.22$, $p = .02$; hyperactivity, $r(114) = -.24$, $p = .01$; and the ADHD composite, $r(114)= -.23$, $p = .01$). Planning times, the proportion slope variable, and the Planning Time × Accuracy score did not correlate with these indices. Backward digit span was negatively related to the inattention index, $r(114) = -.20$, $p = .03$, and the ADHD composite, $r(114) = -.21$, $p = .03$. Partial correlations controlling for backward span were still significant in the associations between the number of perfect solutions and the hyperactivity index, $r\ (113) = -.21$, $p = .03$, and the ADHD composite, $r(113) = -.18$, $p = .05$ but not the inattention index.

Go-No-Go hit rates were associated with inattention, $r\ (111) = -.20$, $p = .03$; hyperactivity, $r\ (111) = -.22$, $p = .02$, and the ADHD composite, $r(111) = -.24$, $p = .01$. Go-No-Go false alarm rates were also associated with inattention, $r\ (111) = .34$, $p < .001$, and the composite, $r(111) = .29$, $p = .001$, but not with hyperactivity, $r\ (111) = .08$, ns. Target detection hit and false alarm rates were not associated with CRS scores. Partial correlations between TOL perfect solutions and the CRS variables remained significant after Go-No-Go hit rates or false alarm rates were controlled.

DISCUSSION

Several inferences can be made regarding how planning improves in adolescence, how it is impacted by concurrently maturing cognitive skills, and its association with self-reported impulsivity. Similar to Asato et al. (2006), age-related improvements were observed on most problem sets and in perfect task solutions until ages 15–17. Others have reported improved accuracy into the 20s (DeLuca et al., 2003; Huizinga et al., 2006). Results for planning times were more complex. Longer initiation times could reflect planning or difficulties with task understanding. Shorter initiation times could reflect impulsivity, or they could reflect increased processing speed. Initiation times for easier versus more challenging problems appeared to reflect distinct processes and/or task approaches, which is not surprising given that perceptual and motor skills may facilitate success on easier problems. On three-move problems, 9–11-year-olds performed more slowly than 12–17-year-olds, which may be due to the fact that these trials come early in the trial sequence, while on more difficult problems they performed significantly faster than 15–20-year olds but with less accuracy. This speed–accuracy trade-off under more challenging conditions may reflect the “act before thinking” style that characterizes poor planning. More difficult problems may selectively capture this compromise.

A distinction has been made between the mental preplanning that occurs before problem initiation and the online planning that ensues with execution (Goel & Grafman, 1995; Ward & Allport, 1997). Because the CANTAB’s TOL variant does not limit completion times or the number of moves to execute each problem, it may be well suited to isolate individual differences in preplanning. The significant positive but moderate correlation observed between perfect solutions and average planning times indicates that deliberation before beginning a problem supports accurate performance only in some cases or instances. Increased initiation times did not necessarily imply that subsequent execution times would be shorter, verifying that the pre- versus online planning task phases may not be inversely correlated (Phillips, Wynn, McPherson, & Gilhooly, 2000). Our finding of a moderate positive correlation between planning and execution times seems to support that deliberate planning extends throughout the task. Yet, increased planning times contributed to more accurate performance while increased execution times contributed to poor performance. The implications are threefold. First, increased deliberation on more difficult problems seems to reflect the type of cognitive preplanning that one hopes to assess via the TOL. Second, the optimal performance pattern may be for planning and execution times to be inversely correlated. Third, the interpretation of the TOL's timing variables must be considered relative to performance quality. To the extent that better performance is associated with more deliberation, the pre-planning processes demanded by the task continue to mature until ages 15–17. When the proportion of total solution time devoted to planning was examined, 15–20-year-olds devoted increasing amounts of time to planning versus execution as problem complexity increased as opposed to 9–14-year-olds. Thus, the modulation of planning skills according to task demands appears to be actively maturing in mid-adolescence.

These data also support relations between planning and other executive functions, such as WM (digit span) and response control (Go-No-Go performance). Accurate TOL performance was associated with these processes as evidenced by significant first-order correlations, consistent with other studies (Asato et al., 2006; Huizinga et al., 2006; Welsh et al., 1999). Since these processes are concurrently maturing, a question of interest concerns whether any of them contribute to performance independent of the others. Regression analyses revealed that only digit span backward contributed to TOL accuracy (perfect solutions) after age and IQ (also significant predictors) were controlled. The selectivity of this relationship coheres with reports emphasizing the importance of WM for planning (Gilhooly et al., 2002; Phillips, Wynn, Gilhooly, Della Sala & Logie, 1999; Welsh et al., 1999). In addition, higher scores on the Accuracy-by-Planning Time index were predicted only by digit span backward, suggesting that strong WM skills support the ability to use mental preplanning adaptively regardless of age or intellect. Further exploration of the variance shared by digit span backward and the TOL may clarify which processes and neural structures facilitate planning in adolescents.

Despite reported associations between TOL performance and IQ (Unterrainer et al., 2004; Zook et al., 2004), correlations in this sample, which was characterized by above-average full-scale IQs, were significant but small in magnitude. With the exception of Huizinga et al. (2006), studies that have estimated IQ through matrix reasoning tasks (Unterrainer et al., 2004; Zook et al., 2004) have yielded more robust relations. In addition, Unterrainer et al. (2004) used a TOL variant that extended to seven moves and used a weighted scoring scheme. In Zook et al.'s (2006) TOL task, the number of pegs varied from three to five instead of remaining constant throughout the task. Thus, the IQ/performance relationship may be augmented when fluid intelligence is specifically examined or when the task is made more challenging or unpredictable.

Similarly, our Go-No-Go findings illuminate how IC relates to planning. Others (Huizinga et al., 2006; Welsh et al., 1999) have reported significant TOL/Stroop associations. The Stroop measures interference suppression as opposed to motor inhibition. The proportion of time spent planning versus executing increased here as trials became more complex and the rate of increase across problem sets related to IC independent of age, IQ, and WM. Thus, on-task behavioral control may be disrupted when a person has a weak capacity for sustained response control.

Finally, we observed that CRS inattention and hyperactivity scores were higher in individuals who achieved fewer perfect TOL solutions. The hyperactivity association remained significant when digit span backward performance was controlled, but the attention association did not, suggesting that associations between TOL accuracy and inattention were influenced by WM. In contrast, inhibitory control of motor responses did not mediate the associations between deliberative planning and CRS inattention or hyperactivity scores. Thus, the TOL appeared to be sensitive to individual differences in both task-demonstrated and self-reported motor control, as well as to WM-mediated variations in attentional control. Because the CRS is age-scaled, levels of self-reported inattention and hyperactivity likely measure trait variance. Thus, competent planning may be impacted independently by maturation and by early-emerging individual differences in trait impulsivity.

Limitations of this study include the use of a convenience sample that was not fully representative of the larger population of adolescents due to our selection criteria and included participants who were, on average, relatively high in socioeconomic status with above-average IQ scores. The reported developmental trends represent those that occur in the absence of risk. Because the study excluded clinical conditions, the range of measurement of some variables (e.g., CRS) was restricted, and the magnitude of TOL correlations may have been attenuated. Only three individuals obtained CRS T-scores at or above 60 on the ADHD index, and their TOL scores were not remarkable for their ages; accordingly, the clinical utility of the CRS/TOL association remains to be determined. Also, many of the reported correlations, while statistically significant, were small in magnitude, raising the possibility of Type 1 errors. However, the significant correlations formed a meaningful set of associations predicted a priori and that replicate past studies. Alternatively, the lack of shared method variance across measures may have reduced the magnitude of associations (Campbell & Fiske, 1959). For example, the correlations between the TOL and the CRS involve variables derived from a timed computerized task using highly abstract stimuli and rules in relation to variables derived from a questionnaire that requires respondents to reflect on their real-world behavior and rate it in terms of general tendencies over time and in relation to others. Obviously the core processes involved are different (i.e., visuospatial reasoning vs. self-reflection). Moreover, complex tasks like the TOL inevitably have multi-factorial influences over performance, while self-report inventories are constructed to provide homogenous measurement scales. As a result, correlations between experimental tasks and self-report measures are often small in magnitude (absolute values of .2 to .3), especially when measured in large samples (Mischel, 1968). The scientific and practical significance of small effects measured in large samples can be as great as for any other effect size (Spencer, 1995). It also may be that the small external correlations reflect the TOL's psychometric deficiencies (Bishop, Aamodt-Leeper, Creswell, McGurck, & Skuse, 2001) or simply indicate that performance is influenced by contributions from as-yet-unidentified factors. This issue merits investigation given its implications for understanding the complexity of planning behavior and how it develops actively during adolescence, given the moderate-to-large independent effects of age on TOL performance that emerged in multiple regression analyses.

In conclusion, the inter-correlations observed between TOL performance, IQ, Go-No-Go response inhibition, digit span performance, and CRS scores suggest that meaningful insights were generated regarding adolescent development of planning abilities, but also that influential cognitive processes and individual difference factors remain to be identified. We suggest that a wide range of potential external correlates should be measured in future studies and that TOL problems should include more difficult problems (Raizner & Song, 2002), thereby providing more precise assessment of adolescent developmental improvements in planning within the larger context of other co-developing domains of executive function.

Acknowledgments

The authors thank the University of Minnesota's Center for Neurobehavioral Development, which provided space for data collection.

This project was supported by the National Institute on Drug Abuse, grant R01DA017843, awarded to Monica Luciana.

REFERENCES

Asato MR, Sweeney JA, Luna B. Cognitive processes in the development of TOL performance. Neuropsychologia. 2006; 44:2259–2269. [PubMed: 16797612]

Badcock JC, Michiel PT, Rock D. Spatial working memory and planning ability: Contrasts between schizophrenia and bipolar I disorder. Cortex. 2005; 41:753–763. [PubMed: 16350658]

Berg WK, Byrd DL. The Tower of London spatial problem-solving task: Enhancing clinical and research implementation. Journal of Clinical and Experimental Neuropsychology. 2002; 24:586–604. [PubMed: 12187443]

Bishop DVM, Aamodt-Leeper G, Creswell C, McGurk R, Skuse DH. Individual differences in cognitive planning on the Tower of Hanoi task: Neuropsychological maturity or measurement error? Journal of Child Psychology and Psychiatry. 2001; 42:551–556. [PubMed: 11383971]

Braver TS, Barch DM, Gray JR, Molfese DL, Snyder A. Anterior cingulate cortex and response conflict: Effects of frequency, inhibition and errors. Cerebral Cortex. 2001; 11:825–836. [PubMed: 11532888]

Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin. 1959; 56:81–105. [PubMed: 13634291]

Conners, CK. Conners' rating scales-revised technical manual. North Tonawanda, New York: Multi Health Systems; 2000.

De Luca CR, Wood SJ, Anderson V, Buchanan JA, Proffitt TM, Mahony K, et al. Normative data from the CANTAB I: Development of executive function over the lifespan. Journal of Clinical and Experimental Neuropsychology. 2003; 25:242–254. [PubMed: 12754681]

Gilhooly KJ, Wynn V, Phillips LH, Logie RH, Della Sala S. Visuo-spatial and verbal working memory in the five-disc Tower of London task: An individual differences approach. Thinking and Reasoning. 2002; 8:165–178.

Goel J, Grafman J. Are the frontal lobes implicated in planning functions? Interpreting data from the Tower of Hanoi. Neuropsychologia. 1995; 33:623–642. [PubMed: 7637857]

Huizinga M, Dolan CV, van der Molen MW. Age-related change in executive function: Developmental trends and a latent variable analysis. Neuropsychologia. 2006; 44:2017–2036. [PubMed: 16527316]

Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. Journal of the American Academy of Child and Adolescent Psychiatry. 1997; 36:980–988. [PubMed: 9204677]

Krikorian R, Bartok J, Gay N. Tower of London procedure: A standard method and developmental data. Journal of Clinical Experimental Neuropsychology. 1994; 16:840–850. [PubMed: 7890819]

Luciana M, Conklin HM, Hooper CJ, Yarger RS. The development of nonverbal working memory and executive control processes in adolescents. Child Development. 2005; 76:697–712. [PubMed: 15892787]

Luciana M, Nelson CA. The functional emergence of prefrontally guided working memory systems in 4-to-8 year-old children. Neuropsychologia. 1998; 36:273–293. [PubMed: 9622192]

Luciana M, Nelson CA. Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated Battery: Performance in 4- to 12-year-old children. Developmental Neuropsychology. 2002; 22:595–624. [PubMed: 12661972]

Luna B, Garver KE, Urban TA, Lazar NA, Sweeney JA. Maturation of cognitive processes from late childhood to adulthood. Child Development. 2004; 75:1357–1372. [PubMed: 15369519]

Mischel, W. Personality and assessment. New York: John Wiley; 1968.

Owen AM, Downes JJ, Sahakian BJ, Polkey CE, Robbins TW. Planning and spatial working memory following frontal lobe lesions in man. Neuropsychologia. 1990; 28:1021–1034. [PubMed: 2267054]

Phillips LH, Kliegel M, Martin M. Age and planning tasks: The influence of ecological validity. International Journal of Aging & Human Development. 2006; 62:175–184. [PubMed: 16541929]

Phillips LH, Wynn VE, Gilhooly KJ, Della Sala S, Logie LH. The role of memory in the Tower of London task. Memory. 1999; 7:209–231. [PubMed: 10645380]

Phillips LH, Wynn VE, McPherson S, Gilhooly KJ. Mental planning and the Tower of London task. The Quarterly Journal of Experimental Psychology A. 2000; 54:579–598.

Purcell R, Maruff P, Kyrios M, Pantelis C. Cognitive deficits in obsessive compulsive disorder on tests of fronto-striatal function. Biological Psychiatry. 1998; 43:348–357. [PubMed: 9513750]

Raizner RD, Song J. Raising the ceiling: The Tower of London-Extended version. Developmental Neuropsychology. 2002; 21:1–14. [PubMed: 12058833]

Rhodes SM, Coghill DR, Matthews K. Neuropsychological functioning in stimulant-naive boys with hyperkinetic disorder. Psychological Medicine. 2005; 35:1109–1120. [PubMed: 16116937]

Sahakian BJ, Owen AM. Computerized assessment in neuropsychiatry using CANTAB: Discussion paper. Journal of the Royal Society of Medicine. 1992; 85:399–402. [PubMed: 1629849]

Shallice T. Specific impairments of planning. Philosophical Transactions of the Royal Society of London B. 1982; 298:199–209.

Spencer B. Correlations, sample size, and practical significance: A comparison of selected psychological and medical investigations. The Journal of Psychology. 1995; 129(4):469–475.

Unterrainer JM, Rahm B, Kaller CP, Leonhart R, Quiske K, Hoppe-Seyler K, et al. Planning abilities and the Tower of London: Is this task measuring a discrete cognitive function? Journal of Clinical and Experimental Neuropsychology. 2004; 26:846–856. [PubMed: 15370380]

Veale DM, Sahakian BJ, Owen AM, Marks IM. Specific cognitive deficits in tests sensitive to frontal lobe dysfunction in obsessive-compulsive disorder. Psychological Medicine. 1996; 26:1261–1269. [PubMed: 8931172]

Ward G, Allport A. Planning and problem solving using the 5-disc Tower of London task. Quarterly Journal of Experimental Psychology. 1997; 50:49–78.

Wechsler, D. Manual for the Wechsler Intelligence Scale for Children–Third Revision. San Antonio, TX: The Psychological Corporation; 1991.

Wechsler, D. Manual for the Wechsler Adult Intelligence Scale–Third Revision. San Antonio, TX: The Psychological Corporation; 1997a.

Wechsler, D. Manual for the Wechsler Memory Scale–Third Revision. San Antonio, TX: The Psychological Corporation; 1997b.

Wechsler, D. Manual for the Wechsler Abbreviated Scale of Intelligence. San Antonio, TX: The Psychological Corporation; 1999.

Welsh MC, Satterlee-Cartmell T, Stine M. Towers of Hanoi and London: Contribution of working memory and inhibition to performance. Brain and Cognition. 1999; 41:231–242. [PubMed: 10590821]

Zook N, Welsh MC, Ewing V. Performance of healthy, older adults on the Tower of London Revised: Associations with verbal and nonverbal abilities. Neuropsychology Development and Cognition B: Aging Neuropsychology and Cognition. 2006; 13:1–19.

Zook NA, Davalos DB, DeLosh EL, Davis HP. .Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. Brain and Cognition. 2004; 56:286–292. [PubMed: 15522766]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE 1

Participant Demographics

	9–11-yr-olds	12–14-yr-olds	15–17-yr-olds	18–20-yr-olds
N =	92	84	82	65
Age	10.3 + 0.8	13.5 + 0.7	16.7 ± 1.0	19.4 ± 0.8
Gender (#M:#F)	47:45	39:45	41:41	25:40
Full-Scale IQ	119.1 + 13.7	115.1 ± 11.6	114.9 ± 11.4	116.1 ± 9.4
Mother's Education (yrs)	15.9 + 2.0	16.2 ± 1.9	16.0 ± 2.0	15.2 ± 2.9
Father's Education (yrs)	15.9 + 2.8	15.9 ± 2.5	16.2 ± 2.9	16.0 ± 3.1
Family Income (U.S. dollars)	95164.8 + 71632.3	87654.3 + 39765.6	102075.9 ± 85201.5	105250.0 ± 73438.5

Note: Unless otherwise indicated, values represent means and standard deviations. There were no significant differences between groups in any variables except average age within groups.

TABLE 2

Tower of London, Target Detection, Go-No-Go, and Digit Span Performance by Age Group

	9–11 years	**12–14 years**	**15–17 years**	**18–20 years**
Tower of London				
Perfect Solutions	7.8 ± 1.6	8.8 ± 1.6	9.1 ± 1.8	9.8 ± 1.5
Average Moves 2	2.0 ± 0.1	2.0 ± 0.0	2.1 ± 0.3	2.0 ± 0.0
Average Moves 3	3.2 ± 0.4	3.2 ± 0.4	3.1 ± 0.2	3.1 ± 0.2
Average Moves 4	5.7 ± 1.0	5.3 ± 1.0	5.1 ± 0.9	5.0 ± 0.7
Average Moves 5	7.1 ± 1.5	6.6 ± 1.4	6.3 ± 1.2	5.9 ± 1.1
Plan 2	1666.6 ± 1065.7	1624.9 ± 858.5	1612.1 ± 843. 9	1770.2 ± 1079.2
Plan 3	5635.8 ± 4660.1	4361.2 ± 2828.2	3587.0 ± 1905.8	4316.8 ± 2580.5
Plan 4	6972.7 ± 6156.6	6466.7 ± 5146.8	6562.1 ± 4489.1	7460.4 ± 4909.1
Plan 5	5813.3 ± 6882.1	7589.9 ± 6400.5	8954.4 ± 5973.3	9883.8 ± 6092.7
Average Planning Time	5022.1 ± 3943.6	5010.7 ± 2815.5	5178.9 ± 2549.6	5857.8 ± 3093.8
Planning Time Proportion 2-move	.68 ± .10	.66 ± .09	.65 ± .09	.66 ± .09
Planning Time Proportion 3-move	.77 ± .09	.73 ± .11	.76 ± .08	.77 ± .08
Planning Time Proportion 4-move	.66 ± .12	.68 ± .13	.73 ± .11	.74 ± .12
Planning Time Proportion 5-move	.68 ± .11	.73 ± .12	.80 ± .10	.82 ± .11
Slope of Planning Time Proportions	−.01 ± .04	.02 ± .05	.04 ±.04	.05 ± .05
Target Detection				
Hit Rate	0.94 ± 0.09	0.97 ± 0.05	0.98 ± 0.03	0.98 ± 0.03
False Alarm Rate	0.03 ± 0.04	0.01 ± 0.01	0.01 ± 0.01	0.00 ± 0.01
Reaction Time	266.2 ± 73.2	206.2 ± 61.0	151.7 ± 51.7	158.4 ± 51.8
Go-No-Go Task				
Hit Rate	0.79 ± 0.17	0.79 ± 0.21	0.77 ± 0.19	0.81 ± 0.17
False Alarm Rate	0.41 ± 0.14	0.29 ± 0.15	0.19 ± 0.11	0.15 ± 0.10
Reaction Time	194.3 ± 86.9	127.3 ± 63.2	91.7 ± 50.7	91.1 ± 41.6
Digit Span Forward	5.7 ± 1.1	6.3 ± 1.1	6.8 ± 0.9	7.2 ± 0.9
Digit Span Backward	4.3 + 1.0	4.8 + 1.3	5.2 + 1.1	5.6 + 1.2

Note: Average move variables represent the average number of moves to complete two-, three-, four-, and five-move problems, respectively. “Plan” variables represent the average planning times (copying-following initiation times) for the same problem sets. Values represent means plus/minus one standard deviation.