Journal of Cognition and Development

ISSN: 1524-8372 (Print) 1532-7647 (Online) Journal homepage: http://www.tandfonline.com/loi/hjcd20

Validating the Electric Maze Task as a Measure of Planning

Kelly W. Sheppard & Carol L. Cheatham

JOURNAL OF COGNITION AND DEVELOPMENT, 00(00):1–15

ISSN: 1524-8372 print/1532-7647 online
DOI: 10.1080/15248372.2016.1195387

Validating the Electric Maze Task as a Measure of Planning

Kelly W. Sheppard and Carol L. Cheatham
University of North Carolina at Chapel Hill

Abstract

The Electric Maze Task (EMT) is a novel planning task designed to allow flexible testing of planning abilities across a broad age range and to incorporate manipulations to test underlying planning abilities, such as working-memory and inhibitory control skills. The EMT was tested in a group of 63 typically developing 7- to 12-year-olds. Participants completed 4 mazes designed to alter working-memory demands (by increasing the number of steps required from 6 to 8) and inhibitory control demands (by using a modified Dimensional Change Card Sort task) and 3 standardized measures using the Cambridge Neuropsychological Test Automated Battery. The EMT was found to correlate with measures of visual memory, working memory, and planning. The 6-step mazes were simpler for participants to solve and mapped on to performance on the visual memory task. The 8-step mazes were more difficult and mapped on to performance on the spatial working-memory and planning tasks. Children who were 10 to 12 years old were also better than 7- to 9-year-olds at solving all mazes, as evidenced by fewer errors and fewer errors later in solving the mazes. Younger children also struggled more than older children after a rule switch. Performance on the maze task differentiated 7- to 12-year-old children with better planning skills, and manipulations of the maze task were successful in altering the working-memory and inhibitory control demands.

INTRODUCTION

Planning is an executive function characterized by assessing a goal, determining how to reach the goal, executing the steps, and then evaluating errors and goal attainment (Lezak, 1982). An important attribute (Welsh & Pennington, 1988) that develops from early childhood through adulthood (Luciana, Collins, Olson, & Schissel, 2009; Welsh, Pennington, & Groisser, 1991), planning requires other executive capacities, such as working memory, inhibitory control, and sustained attention (Lezak, 1982; Welsh et al., 1991; Welsh, Satterlee-Cartmell, & Stine, 1999). Fully developed planning requires the ability to remain "outside" a problem to consider multiple possible solutions and the steps needed to reach any given solution (Kaller, Rahm, Spreer, Mader, & Unterrainer, 2008; Lezak, 1982). Currently, the Tower of London (TOL) and the Tower of Hanoi (TOH) are widely used to assess planning. These tasks consist of two sets of discs arranged on three

Correspondence should be sent to Carol L. Cheatham, Nutrition Research Institute, University of North Carolina at Chapel Hill, 500 Laureate Way Rm 1101, Kannapolis, NC 28081. E-mail: carol_cheatham@unc.edu

to five dowels. The stated end goal is one set of discs should match the other set of discs. To successfully match the two patterns of discs, participants must first work out a general plan and then execute the planned moves while keeping the plan and goal in mind.

The TOL and TOH tasks offer good psychometric validity (Kaller, Unterrainer, & Stahl, 2012) but are somewhat rigidly designed. They do not provide a way to systematically test the contributions of underlying executive capacities within the task, and there is not an explicit way to examine the role of stepping "outside" the problem. The TOL and TOH tasks are often discussed interchangeably, but Bull, Espy, and Senn (2004) demonstrated that performance on the two tasks differs across the preschool period. They attributed the differential performance to subtle differences in the directions and goals of the two tower versions. The instructional differences are thought to alter the extent to which children consider the problem in its entirety before attempting a solution (a key element of planning) and which other executive capacities are needed to solve the problem.

Moreover, the data regarding the cognitive abilities that are employed during the TOL and TOH tasks are mixed. Zook, Davalos, Delosh, and Davis (2004) demonstrated that performance on working-memory and inhibitory control tasks predicted performance on the TOH, but only fluid intelligence was related to performance on the TOL in a sample of undergraduate students. Conversely, Welsh et al. (1999) found that performance on inhibitory control tasks was related to TOL performance in undergraduate students and was only weakly related to success on the TOH. In adults, the nonshared variance between the two tasks does not seem to be explained by differences in administration alone (Welsh, Revilla, Strongin, & Kepler, 2000). Thus, there is a clear need for a planning task that allows more explicit and systematic alteration of task demands to facilitate an understanding of the development of planning and other pertinent executive skills.

Maze tasks offer another way to assess planning that is more flexible than the tower tasks (Gardner & Rogoff, 1990). They have been reliably used with children as young as 3 years old (Gardner & Rogoff, 1990; Miyata, Itakura, & Fujita, 2009) and have allowed researchers to compare performance in young children to performance in primates (Volter & Call, 2014) and other animals (Miyata et al., 2009). This type of task also demonstrates the importance of the number of turns and traps in children's performance, how those manipulations can be used to alter the planning requirements, and the different levels of planning (advanced, trial-and-error, and improvisational) children may use in solving the mazes (Gardner & Rogoff, 1990; Volter & Call, 2014). Some of the issues identified by researchers in the use of maze tasks include the need to constrain children's natural inclination to explore (Miyata et al., 2009) and the need to accurately code children's intentions and processes (Gardner & Rogoff, 1990).

We developed the Electric Maze Task (EMT) to address some of these issues in assessing planning abilities. The EMT may also offer a few other benefits: Active participation may counter potential fatigue; near-infrared spectroscopy systems can easily be used simultaneously to collect cortical imaging data; and path versatility affords a variety of mazes for repeated-measures designs. Our goals were a) to develop an easily modifiable task that would assess planning and the contributions of underlying executive capacities; b) to develop a planning task that could be used across a broad age range; and c) to develop a flexible task with which to assess several unique aspects of planning, such as remaining outside the problem.

METHODS

Participants

Sixty-three typically developing 7- to 12-year-olds (M_{age} = 9;7) from the Southeast United States participated in a larger study about the effects of fatty acids on executive functions in middle childhood. The sample was 52% female and 10% Hispanic. Eighty-eight percent identified as Caucasian, 7% as African American, and 3% as Asian American. Participants were excluded if they had diagnoses of developmental delays or any metabolic condition that resulted in a severely restricted diet. Participants came from households with a diverse range of incomes (58% below $90,000), and 66% of households had at least one parent with a college degree.

Behavioral Testing

Participants visited the lab for a 2-hr session during which they were tested using the new EMT. Memory, working memory, and planning were also assessed using the standardized Cambridge Neuropsychological Test Automated Battery (CANTAB).

The Electric Maze Task

Test Materials. The EMT consists of a mat with a 6 × 8 grid of squares and a control box (Figure 1). The control box is connected to beep sensors under each square. Placing a peg in the corresponding hole in the control box turns off the sensor/beep such that the researcher can set a correct path through the maze: Squares without corresponding pegs (incorrect squares) beep when a participant steps on them. For the present study, a collection of red and yellow stars and circles was also made with laminated cardstock and placed on the squares with Velcro© (Figure 1). Blue outline squares made from laminated cardstock were used to mark the beginning and end of the maze. Participants were videotaped as they completed the maze, and the videos were analyzed using Noldus EthoVision 8.5 (Noldus, Spink, & Tegelenbosch, 2001) to provide accurate error, duration, latency, and location information for each maze and participant.

Procedures. Participants were first introduced to the EMT with a short practice maze. Two rules applied to all mazes: Only adjacent squares or immediately diagonal squares are a correct choice (i.e., no jumping over squares), and after an error (i.e., stepping on a square that beeped), one must return to the beginning. The researcher demonstrated these rules and then allowed the participant to complete the practice maze. This practice component ensured that errors could more reliably be attributed to the cognitive demands of the mazes rather than confusion over the rules. Participants were required to complete four steps on the practice mazes.

Participants then completed four mazes: two requiring six steps and two requiring eight steps. A modified version of the Dimensional Change Card Sort (DCCS; Zelazo & Frye, 1998) task was used to systematically alter the inhibitory control requirements of the task. In a traditional DCCS task, children are asked to sort cards that have pictures with two dimensions (typically shape and color). They are first asked to sort the cards on one

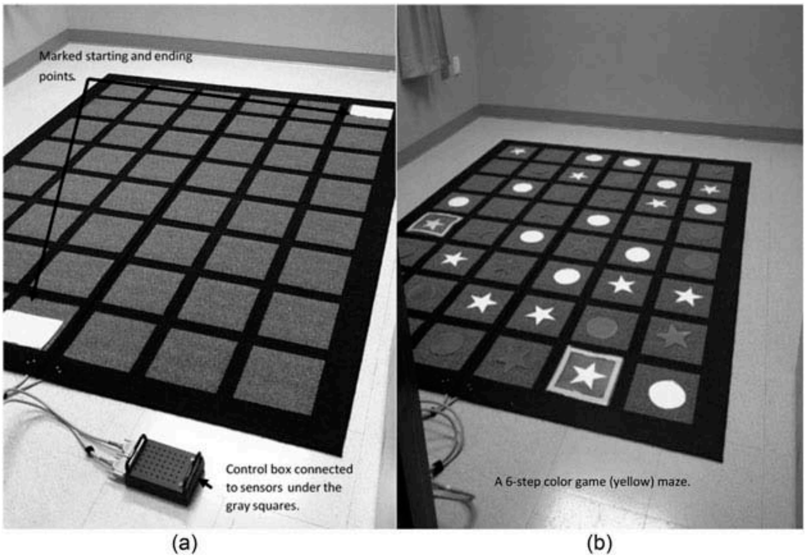

FIGURE 1. The Electric Maze Task (A) and the Electric Maze Task with a modified Dimensional Change Card Sort task (B).

dimension, such as sorting by color, and then they are asked to sort the cards using the other dimension, such as sorting by shape (Zelazo & Frye, 1998). This switch requires inhibiting the old set of rules to successfully complete the task. On the EMT, red and yellow stars and circles were placed on all 48 squares seemingly randomly distributed (Figure 1), but in fact, they were placed to create specific mazes for playing the Color Game and specific mazes for playing the Shape Game. In the Color Game, only one color (either red or yellow) was correct for the whole maze. In the Shape Game, only one shape (either stars or circles) was correct for the whole maze. In all cases, the participant was instructed to stand on the starting square before the researcher began the trial. This removed any uncertainty as to which square was the starting square and which was the ending square. Participants completed a six-step maze and an eight-step maze in each mode, in turn (see Table 1 for an example). Participants were not told how many steps were required to solve each maze. There was only one correct path through any given maze, but all mazes had other possible paths that would have resulted in errors. All mazes were designed to require a counterintuitive move away from the ending square. Each maze required the same number of turns, and the same number of errors was possible. The starting condition and starting position were counterbalanced across participants, and an analysis of the different mazes demonstrated no differences in performance related to the starting mode (color or shape) or starting position (data not shown).

TABLE 1
The script and an example procedure for the electric maze

Procedure before the practice maze: *Participant is led into the room with the maze and told to stand at one end so she or he can see the entire maze. The starting and ending squares for the practice maze are already marked. The researcher goes and stands in front of the starting square for the practice maze*. "This is our maze. Your goal is to go from one marked squared (*point at starting square*) to the other marked square (*point at ending square*) by figuring out the correct squares to step on in between. So I will start here (*researcher steps on maze*), and I can choose any square that is next to the one I am on, like this one, this one, this one, this one, or this one (*indicates each of the five options with foot, does not put weight on them*). I can't jump or skip squares or choose one far away (*mimes skipping a nearby square for one farther away*). I will choose this one (*always the same one, always an incorrect square*). Oh no! The beeping means I have not chosen the correct square. I will step back and try again (*researcher steps back to starting square*). I will try this one next (*chooses the correct square*). Great, no beeping sound means I have chosen the correct square. Now I can choose any square next to the one I am on (*indicates each of the eight options with foot*). I will choose this one (*always the same square, always an incorrect square*). Oh! That is not the correct square, I will go back to the beginning and start again (*moves back to the beginning*). That is the maze. Do you have any questions? *Answer any questions. Have the participant stand on the starting maze, check that the computer is ready to record, and tell the participant to start. Correct any rule errors immediately during the practice maze*.

Maze order was counterbalanced. Procedure for first two mazes: *Turn off beeping while moving pegs to set up each new maze*. Great job, now I will set up your next maze (*move the starting and ending squares for the participant's first maze*). For this maze, we're going to play a game. You can see that there are red and yellow stars and circles all over the maze. We're going to play the Color/Shape Game. In the Color/Shape Game, only one color/shape will be correct for the whole maze. Which do you think will be correct? (C*hild either answers because they can see that the starting and ending squares are the same or the researcher will help them see they are the same*). Great, remember not to jump squares and to go back to the beginning if you choose the wrong square.

Procedure for second two mazes: *After setting up the next maze as described above*, Now we're going to play a different game. We're not going to play the Color/Shape Game anymore. We're going to play the Shape/Color Game. In the Shape/Color Game, only one shape/color will be correct for the whole maze. Which do you think will be correct? (*Researcher makes sure the child knows based on the starting and ending squares*.) Great, remember not to jump squares and to go back to the beginning if you choose the wrong square.

Maze	Number of steps	Condition	Rules
Practice maze	4	—	No jumping squares, go back to the beginning after an error
Maze 1	6	Color	Only squares with yellow will be correct, no jumping squares, go back to the beginning after an error
Maze 2	8	Color	Only squares with red will be correct, no jumping squares, go back after an error
Maze 3	6	Shape	Only squares with stars will be correct, no jumping squares, go back after an error
Maze 4	8	Shape	Only squares with circles will be correct, no jumping squares, go back after an error

Standardized Measures

Test Materials

The CANTAB system consists of a touchscreen computer equipped with software that automatically stores data as participants complete the tasks. Participants completed the Spatial Working Memory (SWM) task, the Stockings of Cambridge (SOC) planning task, and the Paired Associates Learning (PAL) memory task. These CANTAB tests have been used reliably with children aged 7 to 12 years old (Luciana, 2003) to test planning and working memory, and here they served to validate the EMT as a similar test.

Procedures

Participants were seated in front of the touchscreen computer, and all tests were administered in the same order. The first task was a baseline motor screening test that was used to ensure participants understood the touchscreen setup. Additionally, a reaction time task was administered to capture baseline reaction and movement time. The tasks used in analyses were then administered as follows.

Stockings of Cambridge. The SOC task is a spatial planning task analogous to the TOL task. The participant is presented with two sets of colored balls in stockings on the screen. The participant must execute the minimum number of moves necessary to make the bottom set of balls match the top set of balls. The computer records the participant's moves in the first trial and presents those same steps in a second trial. The participant must follow the computer by repeating the moves from Trial 1. This process provides the computer with a way to isolate the time required to execute the movements because presumably, no planning or thinking is required in this imitation sequence (second trial).

Spatial Working Memory. The SWM task is a working-memory and planning task that incorporates a heuristic strategy. The participant is presented with boxes on the screen behind which blue tokens are hidden. The computer reveals the contents of a box when the participant chooses it. The goal is to locate all the hidden tokens. Participants must hold in working memory where they have already found a token to avoid repetitive searches. Having a search strategy also reduces the time spent searching and the number of errors made.

Paired Associates Learning. The PAL task is a visual memory task in which participants are shown a set of boxes on the screen that open and close in random order to reveal unique patterns. After all the patterns have been revealed and rehidden, participants are shown each pattern in turn and must choose the box under which the specific pattern was hidden.

Statistical Analyses

Correlations and paired t tests were completed with SAS Version 9.4 software (SAS Institute Inc.), and exploratory and confirmatory factor analyses were completed with MPlus Version 7.11 (Muthén & Muthén, 1998–2012). False discovery rates (FDRs) were used to correct p values for multiple comparisons in the correlation tables using PROC MULTTEST in SAS Version 9.4. The CANTAB outcome variables were calculated by the software and included error, duration, and latency scores. The outcome measures for the EMT were obtained using Noldus EthoVision 8.5. The Noldus experiment was set up to include measures of errors, duration spent on the correct maze path and in the error zones, and latency to errors and to completion. Raw scores from EthoVision on the two six-step mazes were averaged together, and raw scores on the two eight-step mazes were averaged together. Scores on the mazes from the first condition (color or shape) were averaged, and scores on the mazes from the second condition (color or shape) were averaged. Change scores for step number and change scores for condition were then calculated by subtracting the average scores (second condition minus first condition). In this way, we compared changes in performance for six-step and eight-step mazes (increasing working-memory demands) and after a rule switch was required (increasing inhibitory control demands). Supplementary Table S1 contains a description of all maze and standardized measure variables.

RESULTS AND DISCUSSION

Reliability

Data were visually and statistically assessed for violations of assumptions of normality, equality of error variance, independence, and linearity. No violations were found, and no observations were found to unduly influence results. The maze variables were reliable across all participants (Cronbach's α = .88) and when divided by age group (7- to 9-year-olds, M_{age} = 8;2, Cronbach's α = .87; 10- to 12-year-olds, M_{age} = 11;0, Cronbach's α = .82). A confirmatory factor analysis demonstrated that the six- and eight-step mazes factored separately, $\chi^2(127) = 283.967$, $p < .001$, Comparative Fit Index = .92, Tucker-Lewis Index = .9, root mean square error of approximation = .14, standardized root mean square residual = .09. Even after separating the six- and eight-step mazes, the six-step maze variables were reliable for all participants (Cronbach's α = .89), 7- to 9-year-olds (Cronbach's α = .89), and 10- to 12-year-olds (Cronbach's α = .86), and the eight-step mazes were reliable for all participants (Cronbach's α = .88), 7- to 9-year-olds (Cronbach's α = .88), and 10- to 12-year-olds (Cronbach's α = .88).

Associations With Standardized Measures: Six- and Eight-Step Mazes

Task validity was supported by the pattern of correlations between the EMT scores and standardized measures. Outcomes on the six-step mazes correlated most with outcomes from the PAL memory task, particularly the memory score, the number of stages completed on the first try, and the total errors made (Table 2). The PAL memory score is the total number of patterns recalled correctly across all stages, which was negatively correlated with the total errors made on the six-step mazes ($r = -.44$, $p < .01$) and the duration spent in the error zones ($r = -.49$, $p < .001$). The pattern of correlations between the six-step mazes and standardized measures was similar for the 7- to 9-year-olds and 10- to 12-year-olds, indicating that the six-step mazes required similar skills across the whole age range. The 10- to 12-year-olds did perform significantly better than the 7- to 9-year-olds by making fewer errors, avoiding errors later in the maze, and spending less time in the error zone (see Figure 2). Supplementary Table S2 contains all age-group comparisons for all tested maze outcomes. All *p* values are corrected for FDR.

Outcomes on the eight-step mazes correlated consistently with outcomes from the SOC planning task and the SWM working-memory task (Table 3). The number of planning problems solved in the minimum number of moves is a quintessential planning outcome score in that it represents consistent use of optimal planning. A higher number of problems solved without any extra moves, especially as the problems become more difficult, is indicative of planning. The four-move and five-move problems on the SOC task all require making counterintuitive moves, and the more problems participants could solve in the minimum number of moves, the more

TABLE 2
Pearson's correlations between six-step maze outcomes and standardized tests

	Maze errors	*Rule errors*	*Persev errors*	*Latency to last error*	*Duration in error zone*	*PAL mem score*	*PAL first try stages*	*PAL total errors*	*SWM btw errors*	*SOC solved min moves*	*SOC MST five-move*	*SOC mean moves five-move*
Total errors	.3*	.86***	.86***	.78***	.69***	−.44**	−.33*	.3*	.21	−.02	.22	−.02
Maze errors		.06	.12	.15	.14	−.32*	−.28	.3*	.15	−.1	.06	−.03
Rule errors			.74***	.63***	.53***	−.28	−.2	.14	.16	−.11	.33*	−.02
Persev errors				.67***	.66***	−.32*	−.19	.2	.16	.02	.04	−.1
Latency to last error					.82***	−.39**	−.27	.29*	.26	−.08	.23	.12
Duration in error zone						−.49***	−.32*	.37**	.26	−.15	.15	.06
PAL mem score							.77***	−.85***	−.42**	.12	−.12	.04
PAL first try stages								−.58***	−.42**	−.01	−.09	.02
PAL total errors									.41**	−.09	.07	−.1
SWM between errors										−.32*	.11	.26
SOC solved min moves											−.28	−.51***
SOC MST five-move												.4**

*$p < .05$. **$p < .01$. ***$p < .001$. p values have been false discovery rate-corrected to control for multiple comparisons. SOC = Stockings of Cambridge; SWM = Spatial Working Memory; PAL = Paired Associates Learning; persev = perseverative; mem = memory; btw = between; MST = mean subsequent thinking time; maze errors = errors in which the participant broke one of the rules that apply to all conditions; rule errors = errors in which the participant broke a rule of the Color or Shape Game.

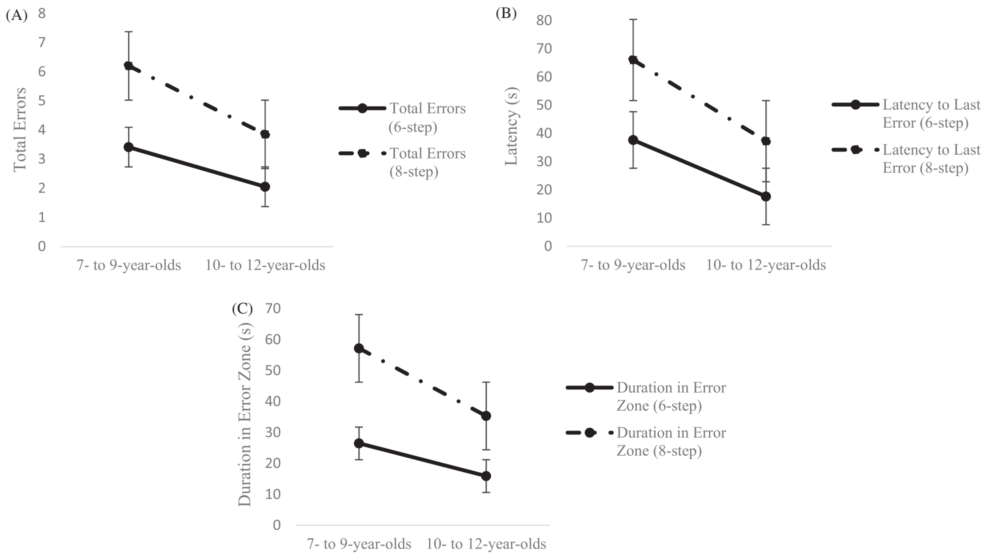

FIGURE 2. Age-related differences in maze performance. (A) Difference in total errors on six- and eight-step mazes. (B) Differences in latency to last error on six- and eight-step mazes. (C) Differences in duration in error zone on six- and eight-step mazes. All differences are statistically significant; t statistics reported include correction for unequal variances between groups when the Folded F test was significant: six-step total errors, $t(61) = 2.03$, $p < .05$, $d = 0.52$; eight-step total errors, $t(55.83) = 3.19$, $p < .05$, $d = 0.85$; six-step duration in error zone, $t(51.25) = 2.06$, $p < .05$, $d = 0.58$; eight-step duration in error zone, $t(56.58) = 2.68$, $p < .05$, $d = 0.67$; six-step latency to last error, $t(42.73) = 2.49$, $p < .05$, $d = 0.62$; eight-step latency to last error, $t(56.13) = 2.71$, $p < .05$, $d = 0.72$.

likely it was that they were successfully mentally stepping outside the problem. The number of problems solved in the minimum number of moves was correlated with the latency to the last error on the eight-step mazes ($r = -.38$, $p < .01$). The measure of latency to the last error is an indication of planning as well, because making errors later in the maze is an indication that the participant has not successfully planned a solution. Indeed, participants who made errors later in the maze were more likely using trial and error to reach the endpoint. Latency to the last error on the eight-step mazes was highly correlated with total errors made on the eight-step mazes ($r = .78$, $p < .001$). The mean moves on the five-move SOC planning problems variable was correlated with perseverative errors on the maze ($r = .45$, $p < .01$). Perseverative errors are an indication of a lack of updating information with the information gained from making an error. In general, the eight-step mazes required working-memory and planning abilities in this age range (7–12 years) and differentiated those who were able to make counterintuitive moves and subsequently update their plans as they acquired new knowledge from those who could not do so. The pattern of correlations between the eight-step mazes and standardized measures was similar for the 7- to 9-year-olds and 10- to 12-year-olds, indicating that the eight-step mazes required similar skills across the whole age range. The 10- to 12-year-olds did perform significantly better than the 7- to 9-year-olds by making fewer errors, avoiding errors later in the maze (indicative of a successful plan), and spending less time in the error zone (see Figure 2). All p values are FDR-corrected.

TABLE 3
Pearson's correlations between eight-step maze outcomes and standardized tests

	Maze errors	*Rule errors*	*Persev errors*	*Latency to last error*	*Duration in error zone*	*PAL mem score*	*PAL first try stages*	*PAL total errors*	*SWM btw errors*	*SOC solved min moves*	*SOC MST five-move*	*SOC mean moves five-move*
Total errors	.29*	.68***	.73***	.78***	.71***	−.19	−.14	.17	.35*	−.41**	.16	.34*
Maze errors		.02	.06	.22	.14	−.18	−.11	.32*	.1	−.07	.09	−.08
Rule errors			.33*	.48***	.41**	.05	.14	−.11	.19	−.25	.01	.19
Persev errors				.53***	.54***	−.08	−.06	.13	.3*	−.39**	.21	.45**
Latency to last error					.87***	−.3*	−.22	.22	.3*	−.37**	.24	.35*
Duration in error zone						−.23	−.18	.17	.29*	−.38**	.28	.39**
PAL mem score							.77***	−.85***	−.42**	.12	−.12	.04
PAL first try stages								−.58***	−.42**	−.01	−.09	.02
PAL total errors									.41**	−.09	.07	−.1
SWM btw errors										−.32*	.11	.26
SOC solved min moves											−.28*	−.51***
SOC MST five-move												.4**

*$p < .05$. **$p < .01$. ***$p < .001$. p values have been false discovery rate-corrected to control for multiple comparisons. SOC = Stockings of Cambridge; SWM = Spatial Working Memory; PAL = Paired Associates Learning; persev = perseverative; mem = memory; btw = between; MST = mean subsequent thinking time; maze errors = errors in which the participant broke one of the rules that apply to all conditions; rule errors = errors in which the participant broke a rule of the Color or Shape Game.

Associations With Standardized Measures: DCCS Manipulation

Data were subjected to paired *t* tests to compare performance after the rule switch. A lack of significant differences in these *t* tests indicated that the manipulation may not have been challenging for this age range. Participants committed more maze errors during the first condition than during the second condition, $t(62) = -2.10$, $p < .05$, Cohen's $d = 0.26$, and moved more slowly during the first condition than during the second condition, $t(61) = 2.60$, $p < .05$, Cohen's $d = 0.33$. Maze errors were coded when a participant made an error that violated one of the common rules (i.e., rules that applied to all mazes regardless of condition). These results indicate that participants improved in their overall mastery of the mazes and became more comfortable solving mazes across time. Because great strides are made in development from the age of 7 to 12 years old, we also performed paired *t* tests by age group. The 7- to 9-year-olds had longer latency to the last error on the second condition than on the first condition, indicating that they did not plan as well after the rule switch occurred, $t(32) = 2.07$, $p < .05$, $d = 0.36$. This decrease in performance could be due to the increased inhibitory control demands the switch placed on them. The 10- to 12-year-olds did not perform any differently after the rule switch (all *p*s > .05). Supplementary Table S3 shows all comparisons in performance after the rule switch.

Across the sample, the change in the latency to the last error after the rule switch was positively correlated with mean subsequent thinking time on the simpler planning tasks ($r = .31$, $p < .05$). The mean subsequent thinking-time measure is similar to the measure of latency to the last error in that it is an indication of how much time the participants spent thinking about how to solve the problem after they had already started to make moves. The fact that these outcomes were correlated for the simpler planning problems (i.e., problems requiring only three moves and no counterintuitive moves) is an indication that the rule switch was most difficult for those whose planning abilities were least developed in that they struggled even with planning problems that did not require counterintuitive moves.

The pattern of correlations for the two age groups were quite different (Table 4), which is logical because the modified DCCS was most difficult for those with the least-developed planning skills. Performance after a rule switch for the 7- to 9-year-olds was related to performance on the SOC planning task. Increases in latency to the last error and time spent in the error zones of the mazes were positively correlated with mean subsequent thinking time on the simplest three-move planning problems ($r = .42$, $p < .05$). Conversely, latency to the last error was negatively correlated with mean subsequent thinking time on the most difficult five-move planning problems ($r = -.43$, $p < .05$). These data indicate that 7- to 9-year-olds who made errors later in the maze after the rule switch, from which we infer difficulties with planning, also required more time after implementing a plan on the simpler problems. However, those 7- to 9-year-olds also took less time after implementing a solution on the most difficult planning problems. These data suggest that 7- to 9-year-olds may have struggled to such a high degree with the five-move problems that they abandoned any planning attempts.

For the 10- to 12-year-olds, performance after the rule switch was associated with the mean moves on the most difficult five-move SOC planning task. Participants who required more moves to solve the five-move planning problems made more total errors ($r = .55$, $p < .05$), spent more time in the error zones ($r = .55$, $p < .05$), and made errors later in solving the mazes ($r = .55$, $p < .05$). Those who made more total errors after the rule switch also had a higher memory score on the PAL memory task ($r = .46$, $p < .05$), indicating that they were able to

TABLE 4
Pearson's correlations with rule switch scores and standardized measures by age group

	Maze errors	*Rule errors*	*Persev errors*	*Latency to last error*	*Duration in error zone*	*PAL mem score*	*PAL stages first try*	*SOC MST three-move*	*SOC MST four-move*	*SOC MST five-move*	*SOC mean moves five-move*
7- to 9-year-olds (*n* = 33)											
Total errors	.17	.83***	.73***	.72***	.63***	.02	.14	.26	.11	−.43*	−.07
Maze errors		.07	−.01	.22	.29	−.02	.14	.24	.21	−.13	.09
Rule errors			.48*	.46*	.45*	.13	.13	.04	−.01	−.32	.03
Persev errors				.75***	.45*	−.03	.07	.35	.22	−.19	.06
Latency to last error					.64***	.05	.16	.42*	.02	−.43*	−.08
Duration in error zone						−.09	.004	.42*	.05	−.41	−.27
PAL mem score							.68***	−.45*	−.31	−.27	.12
PAL stages first try								−.28	−.04	−.16	.11
SOC MST three-move									.48*	.12	.34
SOC MST four-move										.44*	.14
SOC MST five-move											.1
10- to 12-year-olds (*n* = 30)											
Total errors	−.06	.74***	.86***	.71***	.8***	.46*	.26	.12	.15	.32	.55*
Maze errors		.07	−.12	−.05	.11	.27	.39	−.12	.23	.07	−.03
Rule errors			.52*	.48*	.48*	.43	.23	.05	.04	.12	.36
Persev errors				.61**	.64**	.41	.26	−.02	.2	.31	.53*
Latency to last error					.82***	.23	.2	.09	.03	.12	.55*
Duration in error zone						.41	.34	.06	.26	.39	.55*
PAL memory score							.84**	.1	.2	.11	.1
PAL stages first try								.07	.22	.05	.06
SOC MSTthree3-move									−.08	.23	.47*
SOC MST four-move										.72***	.08
SOC MST five-move											.34

*$p < .05$. **$p < .01$. ***$p < .001$. *p* values have been false discovery rate-corrected to control for multiple comparisons. SOC = Stockings of Cambridge; PAL = Paired Associates Learning; MST = mean subsequent thinking time; MIT = mean initial thinking time; persev = perseverative; maze errors = errors in which the participant broke one of the rules that apply to all conditions; rule errors = errors in which the participant broke a rule of the Color or Shape Game.

remember the location of more patterns across the PAL task. If one relies on memory rather than on planning to solve the problems or mazes, she or he would experience issues with inhibiting the old rule after the switch. Indeed, relying solely on memory could also result in a higher number of executed moves on problems requiring counterintuitive moves, as was found in this sample, as the counterintuitive solutions require inhibition of distracting information.

CONCLUSIONS

The EMT provides a reliable measure of planning that works well with 7- to 12-year-olds and correlates well with standardized measures. Better performance on standardized measures was correlated with better performance on the EMT, and expected age-related differences were seen in performance on all maze configurations. From these initial results, it appears that by using the EMT, we can differentiate between older and younger participants and possibly between those who have fully developed planning and those who do not. Performance on the EMT indicated that those who had difficulty with simpler standardized planning problems struggled most with the mazes relative to others and that the rule switch was most challenging for those who presumably, based on the data available, were relying on memory capacity instead of flexibly employing both inhibition and memory skills.

Manipulations of the maze, such as altering the number of steps or implementing our modified DCCS task, can also be used to create tests of the underlying capacities required for planning. With these manipulations, we were able to elicit different responses from 7- to 9-year-olds and 10- to 12-year-olds that correlated in the expected direction with standardized measures. The initial 6- and 8-step mazes discussed here may not have been difficult enough for this age range. The 6-step mazes were basic, and it appears most participants were able to rely on memory to solve the mazes. Future testing with mazes with more steps (10 and 12 steps) and with more complex inhibitory control manipulations (such as three dimensions) will be needed to completely understand how well these mazes work to test planning in this age range.

It would also be appropriate to test the EMT with other maze tasks. The different requirements of these other tasks may offer additional insight into the EMT as a measure of planning. The EMT addresses several concerns with measures previously employed in the field. Participants received immediate feedback for each step (beeping or no beeping), thereby constraining young children's inherent desire to explore the grid. The structured grid and response modality (walking) allows the use of tracking software to reduce subjective evaluations of behavior, but videos could easily be manually coded if tracking software is not available. The participants made full-body movements that were easy to detect, and time was measured accurately and identically for all participants. Active, mobile participation also made the EMT a fun and engaging task. The main benefit of the EMT is it affords the ability to create a developmentally appropriate set of mazes for a longitudinal study. The underlying requirements of the task do not change, but the difficulty of the mazes and the underlying executive skills required can be altered to match the participants' development. The main drawback to the EMT was the space required, but setup could be completed in less than 15 min if a dedicated space is not available. Further testing across childhood and into adulthood will be needed to create a standard approach to using the EMT as a measure of planning.

ACKNOWLEDGMENTS

The validation of the task was presented at the 2015 Society for Research in Child Development Biennial Meeting in Philadelphia, PA. The authors would like to especially thank Sandra Wiebe for lending her expertise in executive functions.

FUNDING

This research was supported by a grant from the Sigma Xi Grants-in-Aid-of-Research Program (A-8387080249), a Department of Psychology & Neuroscience Dissertation Start-Up Grant to the first author from the University of North Carolina, and research funds granted to the second author by the Nutrition Research Institute and the Department of Psychology & Neuroscience at the University of North Carolina.

SUPPLEMENTAL MATERIAL

Supplemental data for this article can be accessed on the publisher's website.

REFERENCES

Bull, R., Espy, K. A., & Senn, T. E. (2004). A comparison of performance on the Towers of London and Hanoi in young children. *Journal of Child Psychology and Psychiatry*, *45*, 743–754. doi:10.1111/jcpp.2004.45.issue-4

Gardner, W., & Rogoff, B. (1990). Children's deliberateness of planning according to task circumstances. *Developmental Psychology*, *26*, 480–487. doi:10.1037/0012-1649.26.3.480

Kaller, C. P., Rahm, B., Spreer, J., Mader, I., & Unterrainer, J. M. (2008). Thinking around the corner: The development of planning abilities. *Brain and Cognition*, *67*, 360–370. doi:10.1016/j.bandc.2008.02.003

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*, 46–53. doi:10.1037/a0025174

Lezak, M. D. (1982). The problem of assessing executive functions. *International Journal of Psychology*, *17*, 281–197. doi:10.1080/00207598208247445

Luciana, M. (2003). Practitioner review: Computerized assessment of neuropsychological function in children: Clinical and research applications of the Cambridge Neuropsychological Testing Automated Battery (CANTAB). *Journal of Child Psychology and Psychiatry*, *44*, 649–663. doi:10.1111/1469-7610.00152

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*, 461–475. doi:10.1080/87565640902964540

Miyata, H., Itakura, S., & Fujita, K. (2009). Planning in human children (Homo sapiens) assessed by maze problems on the touch screen. *Journal of Comparative Psychology*, *123*, 69–78. doi:10.1037/a0012890

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Noldus, L. P. J. J., Spink, A. J., & Tegelenbosch, R. A. J. (2001). EthoVision: A versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods, Instruments, & Computers*, *33*, 398–414. doi:10.3758/BF03195394

Volter, C. J., & Call, J. (2014). Younger apes and human children plan their moves in a maze task. *Cognition*, *130*, 186–203. doi:10.1016/j.cognition.2013.10.007

Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology*, *4*, 199–230. doi:10.1080/87565648809540405

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, *7*, 131–149. doi:10.1080/87565649109540483

Welsh, M. C., Revilla, V., Strongin, D., & Kepler, M. (2000). Towers of Hanoi and London: Is the nonshared variance due to differences in task administration? *Perceptual and Motor Skills*, *90*, 562–572. doi:10.2466/pms.2000.90.2.562

Welsh, M. C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: Contribution of working memory and inhibition to performance. *Brain and Cognition*, *41*, 231–242. doi:10.1006/brcg.1999.1123

Zelazo, P. D., & Frye, D. (1998). Cognitive complexity and control: II. The development of executive function in childhood. *Current Directions in Psychological Science*, *7*, 121–128. doi:10.1111/1467-8721.ep10774761

Zook, N. A., Davalos, D. B., Delosh, E. L., & Davis, H. P. (2004). Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. *Brain and Cognition*, *56*, 286–292. doi:10.1016/j.bandc.2004.07.003