Understanding planning ability measured by the Tower of London

Gonzalez Marin, G.V.; Bouwmeester, S.; Boonstra, A.M.

Published in:
Psychological Assessment

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

Psychological Assessment
2010, Vol. 22, No. 4, 923–934

1040-3590/10/$12.00 DOI: 10.1037/a0020826

Understanding Planning Ability Measured by the Tower of London: An Evaluation of Its Internal Structure by Latent Variable Modeling

Gabriela V. Koppenol-Gonzalez, Samantha Bouwmeester, and A. Marije Boonstra
Erasmus University Rotterdam

The Tower of London (TOL) is a widely used instrument for assessing planning ability. Inhibition and (spatial) working memory are assumed to contribute to performance on the TOL, but findings about the relationship between these cognitive processes are often inconsistent. Moreover, the influence of specific properties of TOL problems on cognitive processes and difficulty level is often not taken into account. Furthermore, it may be expected that several planning strategies can be distinguished that cannot be extracted from the total score. In this study, a factor analysis and a latent class regression analysis were performed to address these issues. The results showed that 4 strategy groups that differed with respect to preplanning time could be distinguished. The effect of problem properties also differed for the 4 groups. Additional analyses showed that the groups differed on average planning performance but that there were no significant differences between inhibition and spatial working memory performance. Finally, it seemed that multiple factors influence performance on the TOL, the most important ones being the score measurements, the problem properties, and strategy use.

Keywords: cognitive processes, internal validity, latent variable modeling, planning, Tower of London

In 1883, the French mathematician Edouard Lucas developed the Tower of Hanoi (TOH, also known as the Tower of Brahma or the End of the World Puzzle), inspired by the legend of a temple in Vietnam: In the beginning of times the monks of the temple were given a stack of 64 golden disks that differed in size and came on one of three diamond poles. Each disk was placed on top of a bigger disk. The monks had to transfer all of the disks one at a time from one pole to another, such that the tower would have the exact same shape. They had the restrictions that a disk could never be placed on top of a smaller one or outside the poles. When the monks would finish their work, the temple would crumble into dust and the world would vanish (Poole, 1994). On the basis of the TOH, Shallice (1982) developed the Tower of London (TOL) to assess higher order problem-solving capacity, specifically executive planning ability. The TOL consists of a board with three pegs differing in length. Instead of disks, the TOL has three balls usually colored blue, red, and green. Two such boards are used: one with the balls arranged in a start position and one with the balls arranged in a goal position. For each TOL problem, the balls of the start position have to be transferred into the goal position, under three restrictions: (a) The balls have to be moved one at a time; (b) they cannot be placed outside the pegs; and (c) a maximum of three balls are allowed to be placed on the tallest peg, a maximum of two on the middle peg, and a maximum of one on the shortest peg. Figure 1 shows an example of a TOL problem.

Both the TOL and the TOH are widely used instruments for measuring planning ability both in clinical settings and for scientific purposes (e.g., Kaller, Unterrainer, Rahm, & Halsband, 2004; Miyake et al., 2000; Owen, 1997; Ward & Allport, 1997; Welsh, Satterlee-Cartmell, & Stine, 1999). In clinical settings, specifically neuropsychological testing, the primary score of the TOL is the total move score (Culbertson & Zillmer, 2001). This is the number of moves beyond the minimum number of moves required to reach the goal position summed over all problems. For instance, if four TOL problems that can be solved in five moves are all solved in nine moves instead, the total move score, 4 (problems) × 4 (moves that exceed the minimum), is 16. This means that a low total move score reflects good planning. The second main score is the total correct score, which is the total number of problems that are solved in the minimum number of moves. In addition, three time scores can be calculated: the initiation time, execution time, and total time. The initiation time, also called *preplanning time* or *first-move time*, is the period between the presentation of the problem (start and goal positions) and the first move (Berg & Byrd, 2002). The execution time is the period between the first move and the last move, and the total time is the sum of the initiation time and the execution time. These additional time scores are used to obtain information about the efficiency of planning. The total move score is assumed to be the most indicative of planning; all other scores are assumed to contribute to a better understanding of this primary score (Culbertson & Zillmer, 2001). In scientific research, the most frequently used scores are the total move score and the total correct score.

Gabriela V. Koppenol-Gonzalez, Samantha Bouwmeester, and A. Marije Boonstra, Department of Psychology, Erasmus University Rotterdam, Rotterdam, the Netherlands.

Gabriela V. Koppenol-Gonzalez is now at Department of Methodology and Statistics, Tilburg University, Tilburg, the Netherlands.

A. Marije Boonstra is now at Department of Psychology and Education, VU University Amsterdam, Amsterdam, the Netherlands.

We thank Peter de Looff for his assistance in collecting the data and Ruth Koppenol for the recruitment in Avans Hogeschool Breda.

Correspondence concerning this article should be addressed to Gabriela V. Koppenol-Gonzalez, Department of Methodology and Statistics, Tilburg University, P.O. Box 90153, NL-5000 LE Tilburg, the Netherlands. E-mail: g.v.gonzalezmarin@uvt.nl

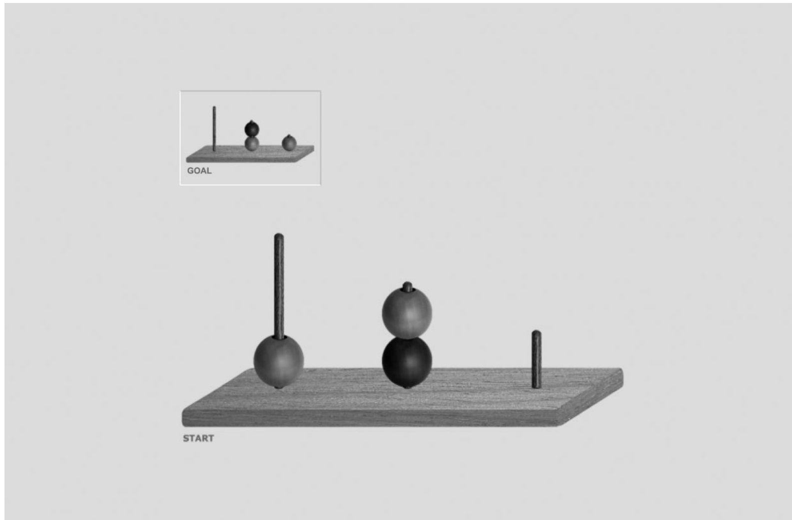

Figure 1. Start and goal positions of a Tower of London problem with a minimum of five moves.

Despite the TOL's frequent use, several problems put the validity of the current interpretation of TOL scores in question. The first and most important problem is that it is not clear how the TOL actually measures planning ability. That is, because of a poor definition of planning and operationalization of the construct, it is rather questionable what the exact relationship is between the theoretical construct planning and TOL scores such as the total move score and the total correct score. This poor definition and operationalization may explain the inconsistency in results of studies that focus on the cognitive processes involved in planning as measured by these total TOL scores (see, e.g., Berg & Byrd, 2002). Planning can be theoretically described as the execution of goal-directed behavior to predict and evaluate outcomes (Kaller et al., 2004) and requires the implementation of a series of steps (subgoals) to achieve the ultimate goal (Owen, 1997). However, it remains unclear which cognitive processes are involved in planning and how these cognitive processes are related to planning as measured by the TOL. Possible cognitive processes that are measured by the TOL seem to be strongly dependent on the general instructions and specific restrictions that are given. For example, when people are allowed to try different possibilities before coming to a solution, a total move score might not reflect much planning. Instead, inhibition may be a more appropriate interpretation of that score. When people presented with the TOL receive the restriction that they have to solve the problems in a minimum number of moves, TOL scores can be assumed to reflect planning in particular. Moreover, it is important to differentiate between the (cognitive) construction of a plan—that is, the actual planning—and the execution of that plan (Goel & Grafman, 1995). This distinction should be reflected in different scores, like a preplanning time score and an execution time score.

A second problem is that deriving a measure of planning ability from the total move and total correct scores, which are the sum of individual problem scores, is not a valid interpretation of these scores. Kafer and Hunter (1997) investigated the relationship between the latent construct planning and total TOL scores, but they could not fit an adequate model. One of their conclusions was that planning may be too complex to be measured with one test. An alternative explanation is that different TOL problems vary in problem properties, such as the number of possibilities for the first move and the minimum number of required moves. These properties may not only influence the difficulty level of a problem but may also require different combinations of cognitive skills. For example, compare a TOL problem that requires a minimum of seven moves and has four possible first moves with a problem that also requires seven moves but has only two possible first moves. In both cases, working memory is an important skill, but in the last case, inhibition may play a less important role than it does in the first case, because the first move has a 50–50 chance of being the correct one. Therefore, participants' performance on different TOL problems cannot be compared. A more valid approach to interpret TOL scores can be achieved by performing a detailed analysis on the problem level to investigate the influence of problem properties on planning performance.

A third problem that questions the validity of TOL score interpretations is that the use of strategies is usually not taken into account, whereas strategy use seems to be an important part of planning (Goel & Grafman, 1995). For instance, Owen, Downes, Sahakian, Polkey, and Robbins (1990) showed that people seem to adopt different kinds of planning strategies when confronted with TOL problems. Specifically, they investigated TOL performance in patients with frontal lobe lesions compared with a control group and measured the mean number of moves above the minimum (mean move score), the total correct score, and the number of problems solved within the maximum number of moves, which was defined as twice the minimum number of moves plus one. Also, they measured the preplanning time and the execution time. There was no significant difference between the groups' preplanning times, but the execution time of the frontal lobe patients was significantly longer than that of the control group. This, together with the finding that frontal lobe patients were less likely to adopt an effective strategy on a working memory task on which they performed worse than the control group, led to the conclusion that frontal lobe patients did not plan efficiently when solving TOL problems. They did, however, solve even the most difficult problems within the maximum number of moves and, therefore, it seemed that frontal lobe patients at least made an effort to come to a solution. The results of this study show that the members of the control group seemed to plan the moves during the preplanning time, whereas the frontal lobe patients seemed to plan the moves after their first move, even though they spent the same amount of time thinking about that first move. Owen et al. (1990) concluded that the frontal lobe patients spent their preplanning time in a less efficient way. In a nonclinical sample, a subgroup of participants may also plan inefficiently during the preplanning time. These participants seem to adopt a trial-and-error strategy (Kovács, 2007). Before making their first move—that is, during preplanning time—they may not be concerned with the actual planning of their moves, but instead they are thinking of their previous erroneous moves so that they can start differently this time. This form of inefficient planning is reflected by a long preplanning time but not necessarily a correct response. Another possibility is that these participants may think they know the first move when a problem is similar to a previous problem and, therefore, they make the same first move. This is then reflected by a short preplanning time but not necessarily a correct response. The study of Owen et al. (1990) also shows that a combination of scores, such as correct scores and time scores, can provide information about the planning strategies people adopt. This is useful information that cannot be obtained using only a total score. Furthermore, the use of a particular strategy should be taken into account, because it may depend on specific problem properties. For example, people with a trial-and-error strategy are expected to solve slightly different kinds of TOL problems in the same way, whereas others may look ahead and adapt their moves depending on the specific problem properties.

It can be concluded that although the TOL is a widely used instrument, several problems undermine the validity of the interpretation of its currently used scores in research. When separate scores are calculated for each TOL problem and problem properties and strategy use are taken into account, the interpretation of TOL scores may be more valid when attempting to measure planning ability. Our aim in this study is to enhance the internal validity of the TOL scores via an in-depth analysis on the TOL-problem level. First, we want to determine whether the cognitive processes hypothesized by the theory and described in detail in the next paragraph can be distinguished empirically. Next, we want to investigate whether the different strategies described by Owen et al. (1990) can be revealed empirically in a nonclinical sample by analyzing the preplanning time on several TOL problems. Our final aim is to determine the effect of problem properties on planning strategies and the relationship between these planning strategies and planning performance.

Planning: Theory and Empirical Evidence

The TOL and the TOH can be described as executive function tasks that measure specifically planning ability. It is generally assumed that working memory and inhibition are the most important components of executive functions that contribute to performance on the TOL and the TOH (Goel & Grafman, 1995; Huizinga, Dolan, & van der Molen, 2006; Miyake et al., 2000). Working memory requires the active manipulation of information (Miyake et al., 2000; Owen, 1997). For the tower tasks, the information to be actively manipulated in working memory concerns the specific moves required to achieve the goal state. Inhibition requires the deliberate suppression of a dominant or automatic response (Miyake et al., 2000). For the tower tasks, the dominant response concerns moving a ball directly into its goal position. This response has to be inhibited to carefully plan all the moves and solve the TOL problems efficiently. Besides working memory and inhibition, shifting between mental sets is also an example of an executive-functions component, but it is not found to be a predictor of TOL and TOH performance (Huizinga et al., 2006; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Miyake et al., 2000).

Because of the apparent similarity between the TOH and the TOL, it seemed obvious that both tasks capture the same cognitive processes (Welsh et al., 1999). However, the results of studies in which these processes were investigated are inconsistent (Berg & Byrd, 2002) and the relationship between the TOH and the TOL has not been empirically demonstrated. For example, Welsh et al. (1999) showed that both working memory and inhibition strongly predict TOL performance, but they also showed that inhibition weakly predicted TOH performance and that working memory did not predict performance on the TOH. Furthermore, 84% of the variance in scores was not shared between the two tower tasks. Zook, Davalos, DeLosh, and Davis (2004) found that 93% of the variance in scores was not shared between the TOL and the TOH and that the scores of both tasks captured different variance in scores of the Matrix Reasoning subtest from the Wechsler Abbreviated Scale of Intelligence. Moreover, working memory and inhibition predicted TOH performance and only inhibition predicted TOL performance. However, in the study of Owen et al. (1990), spatial working memory did play a role in TOL performance.

All of these results taken together not only show that both tasks seem to tap different cognitive processes but also show that it is not clear how the TOL exactly measures planning and which underlying cognitive processes influence performance on the TOL. When studies were compared, it turned out that the administration of the TOL differed with respect to the calculated scores that were used. These scores concerned the total correct score (Owen et al., 1990; Welsh et al., 1999) and the average total move score (Zook et al., 2004); in addition, in only one study, the additional preplanning and execution times were used (Owen et al., 1990). These studies do not use the same TOL scores, but these different scores are all interpreted as measuring planning ability, which makes the validity of such an interpretation questionable. Because the measurement of planning ability strongly depends on the administration of the TOL, it is important that different studies use exactly the same administration method and calculated scores.

Another important issue that threatens the internal validity of TOL scores is that the individual TOL problems are not comparable. TOL problems differ with respect to specific problem properties and these properties may require different cognitive skills. Therefore, performance on the TOL should be analyzed on the problem level.

Several researchers used factor analysis to investigate the extent to which the underlying cognitive processes of planning can be considered separate constructs. Miyake et al. (2000) performed a confirmatory factor analysis on scores of three different executive functions tests and their contribution to performance on global executive tasks. The results showed that inhibition and TOH performance loaded on the same factor. It also turned out that three cognitive processes (shifting of mental sets, updating of working memory representations, and inhibition) were clearly distinguishable but not completely independent constructs. This is an important finding, because it has a major impact on the interpretation of what is being measured when using executive functions tasks. If different TOL problems with particular problem properties require different cognitive skills, it is important to know which problem (property) is related to which skill. Because these three cognitive processes seem to be distinguishable, it should be worthwhile to investigate their role in the different TOL problems. The results of Miyake et al. (2000) were confirmed using the TOL in a child population by means of an exploratory and a confirmatory factor analysis (Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003). In this case, inhibition as measured by the matching familiar figures task and the number of TOL problems solved in the minimum number of moves loaded on the same factor. This relationship can be explained with the notion that the children were allowed to make a maximum number of moves of twice the minimum plus one. This means that they were allowed to try out some possible first moves. As a consequence, it is possible that only the children that could inhibit this "trying out" actually planned their moves first. These children were therefore more likely to solve the problems in the minimum number of moves. Huizinga et al. (2006) also performed a confirmatory factor analysis on tasks that tap inhibition, working memory, and shifting, and they investigated the relationship between performance on these three tasks and performance on executive tasks such as the TOL and the Wisconsin Card Sorting Test. The results showed that inhibition as measured by a Stroop task with pictures instead of words was a significant predictor of perfect solutions only in 21-year-olds and that inhibition was not a significant predictor of the additional moves score and preplanning time. This means that inhibition has an influence when the problems are voluntarily solved in the minimum number of moves, but it does not affect the amount of moves that exceed this minimum or the amount of time needed before making the first move. Working memory and shifting did not seem to have an important influence on TOL performance.

Studies that use the same administration method and calculated scores lead to more consistent findings that support the idea that, in particular, inhibition is related to TOL performance. However, this relation concerns only the TOL problems that are voluntarily solved in the minimum number of moves. None of these aforementioned studies applied a perfect solution approach. Applying this approach means that all participants have to reach the goal position in the minimum number of moves and are not given the opportunity to undo a move. This approach forces participants to plan carefully (Berg & Byrd, 2002) and may not differentiate between participants' inhibition ability, because they are actually instructed to inhibit.

It can be questioned whether planning is measured at all when a perfect solution approach is not applied. According to several researchers, the TOL would be a better instrument for assessing planning if it had the restriction that a move cannot be undone (Goel & Grafman, 1995; Murji & DeLuca, 1998). Otherwise, TOL performance may be more dependent on inhibition than on planning ability, because only participants that inhibit the tendency to move a ball directly into its goal position are able to plan the subsequent moves carefully and solve the problems efficiently. This could explain why inhibition measured by the Stroop is often found to be related to TOL performance when participants are not restricted in the number of moves they are allowed to make. This finding is rather striking, because inhibition as measured by the Stroop (i.e., the rapid inhibition of an automated response) does not seem to be the same type of inhibition that is needed when suppressing the tendency to move a ball directly into its goal position in the TOL. For example, Huizinga et al. (2006) investigated the construct inhibition and found three manifest inhibition variables that seem conceptually similar (measured by the Stroop, Eriksen flankers, and stop-signal tasks) showed low or negative correlation and could not be united into one latent variable, implying the existence of different types of inhibition. It is possible that people with high planning ability also have high inhibition but that the planning needed to solve the TOL in the minimum number of moves is related to a type of inhibition other than that needed to obtain a low interference score on the Stroop.

Although we do not expect inhibition to be related to TOL performance when a perfect solution approach is applied, spatial working memory is expected to influence TOL performance. However, the relationship between working memory and TOL performance is not well established yet. In the aforementioned studies, several tasks were used to measure working memory. All of these tasks specifically require visual working memory: Participants had to recall static visual stimuli or patterns of static visual stimuli. However, TOL problems can be considered dynamic visual stimuli, because they require not only the planning of movements from one peg to another but also the recalling of these movements to execute that plan (i.e., making the moves). Unfortunately, there is no memory task that measures especially this spatial component of working memory. Therefore, we constructed a new task that specifically measures recall of movement.

Problem Properties of the TOL

When using the TOL, even if the same administration method is used, it is likely that the sets of TOL problems are not exactly the same over different studies. The original Shallice TOL, for instance, has 1,260 possible problems (Berg & Byrd, 2002). However, researchers rarely define the problems they use or their reasons for selecting specific problems. This might be problematic, because Kaller et al. (2004) showed that the problem properties have an important influence on the grade of difficulty of TOL problems, because they seem to tap different cognitive processes. Recall the example of solving a TOL problem with a minimum of seven moves and four possible initial moves, which requires both working memory and inhibition, compared with a TOL problem with a minimum of seven moves but only two possible initial moves, in which inhibition plays a less important role. In our study, we distinguish three important properties: minimum number of moves, number of possible initial moves, and alternative goal moves. The solution, or goal position, can be achieved in a minimum number of moves ranging from one to eight depending on the specific start and goal position. The start positions have a specific number of possible initial moves ranging from two to four (Berg & Byrd, 2002). An alternative goal move requires the inhibition of moving a ball immediately into its goal position to avoid blocking subsequent moves (Hodgson, Bajwa, Owen, & Kennard, 2000; Kaller et al., 2004).

Strategies

Besides the influence of problem properties on performance, people may adopt different strategies when confronted with a TOL problem, which may also explain an important source of variation in performance. For instance, some people try to solve the problems by adopting a trial-and-error strategy, whereas others try to look ahead and think through every move before actually making the first one. These strategies influence planning performance, because a specific strategy is more efficient than another and leads to a longer or shorter preplanning time and a higher or lower correct score on a specific TOL problem. However, because strategy use may interact with problem properties, it is important to analyze strategy use on TOL-problem level.

When a perfect solution approach is applied, the preplanning time per item can be expected to give information about the strategies people adopt, whereas a correct score per item can give information about the efficiency of that strategy. Therefore, we expect that at least three strategy groups can be distinguished on the basis of their preplanning time. Table 1 summarizes the hypotheses with respect to the three expected strategy groups. We expect that inhibition will not play a major role in performance or preplanning time, because participants are instructed to inhibit, that is, they have to solve the problems in the minimum number of moves. However, spatial working memory is expected to play a significant role in the execution of the planned movements of TOL problems. Furthermore, we expect three different groups within our sample. The first group is characterized by a short preplanning time and can be described as not adopting an effective strategy. Participants in this group have the tendency to make their first move too fast, which is prone to result in an incorrect response. Therefore, people in this group are expected to show low performance on the TOL. The influence of the problem properties is expected to be small, because these people do not adjust their planning to the problem properties. Therefore, spatial working memory cannot be related to performance on the TOL in this group. The next two groups are expected to show a long preplanning time. Participants in one group can be described as having efficient planning ability, because during preplanning time they are consciously planning their moves. Therefore, the problem properties are expected to influence their preplanning time. Participants in this group are expected to show high performance on the TOL. Moreover, they are expected to show high performance on spatial working memory. Participants in the other group can be described as having inefficient planning ability. They spend a long time thinking before they make their first move, but they spend this time inefficiently. People in this group may not be able to process all of the task requirements and, therefore, they are expected to show low performance on the TOL and low performance on spatial working memory. Moreover, because of inefficient planning, the effect of problem properties on planning is expected to be small.

Research Questions and Hypotheses

For the first research question, we investigated whether one or more cognitive processes explain variation in planning performance on 12 TOL problems that differed with respect to three important properties. Because we applied a perfect solution approach, we did not expect inhibition to explain much variance. Therefore, we hypothesized that one underlying factor explains the variation in TOL-problem performance and that this factor can be interpreted as spatial working memory. For the second research question, we wanted to determine whether strategy groups can be distinguished on the basis of preplanning time. It is hypothesized that the influence of problem properties and working memory differs for the different strategy groups and that performance on the TOL problems also differs for the different strategy groups (see Table 1).

Method

Participants

188 students from 17.6 to 31.1 years of age participated in this study (*M* age = 20.9 years, *SD* = 2.6). The sample included 105 women and 83 men. Within this sample, 154 participants were university students (Erasmus University Rotterdam) and 34 participants were academy students (Avans Hogeschool Breda). Participants received course credit, a small present, or €7.50.

Materials

University students completed three tasks on a computer in a cubicle or separated room in the behavioral lab of Erasmus University Rotterdam. The tasks were the 12 TOL problems, the Stroop to measure inhibition, and the Swom to measure spatial working memory. Academy students completed the tasks on a 15.4-in. Sony Vaio laptop in a separate room in Avans Hogeschool Breda.

The TOL was designed using Adobe Macromedia Flash Player 8.0. The task has the same problem space as that of the Shallice TOL and follows the two recommendations of Berg and Byrd (2002) for computerized versions: Its monitor image is held as similar in appearance to the handout version as possible and the movements necessary to replace the balls are similar to those required in the handout version. This means that the mouse could be used to drag the balls from one peg to another. Table 2 shows

Table 1
Summary of the Hypothesized Strategy Groups

Strategy	Preplanning time	Performance	Influence properties	Inhibition	Spatial working memory
No effective strategy	Short	Low	Small	Average	Unknown
Efficient	Long	High	Large	Average	High
Inefficient	Long	Low	Small	Average	Low

Table 2
Problem Properties of the Tower of London Problems Used in This Study

Problems	Start position	Goal position	Number of moves	Alternative goal move	Number of possible initial moves
Practice	65	13	3	Absent	4
Practice	45	52	4	Absent	4
1	16	36	5	Absent	2
2	24	44	5	Absent	3
3	33	55	5	Absent	4
4	14	36	5	Present	3
5	21	41	6	Absent	2
6	42	61	6	Absent	3
7	25	52	6	Absent	4
8	13	32	6	Present	4
9	11	41	7	Absent	2
10	34	64	7	Absent	3
11	25	54	7	Absent	4
12	64	42	7	Present	3

Note. For the interpretation of the numbering system of start and goal positions, see Berg and Byrd (2002).

the 12 TOL problems, their start and goal positions, and the structural properties. The task consisted of two practice trials that required three and four moves, four possible initial moves, and no alternative goal move. The 12 experimental trials can be divided according to number of moves: The first block consists of four trials that require five moves, the second block consists of four trails that require six moves, and the last block consists of four trials that require seven moves. The problems were randomized within each block, so participants always started with four 5-move problems and always ended with four 7-move problems. Response time and correct score were measured for each problem. Because participants were drawn from a nonclinical sample and the age range was rather small, they were assumed not to differ in motor ability.

The Stroop task was transformed into a computerized version using E-Prime 1.1 for the regular computers and E-Prime 2.0 Beta Program for the laptops. The task was programmed to have the same appearance and procedure as the Dutch handout version of the Stroop Color–Word Test (Hammes, 1978). In the manual for this Dutch version of the Stroop task, high test–retest correlations of the three time scores are reported: .81 for the word naming times, .87 for the color naming times, and .89 for the interference times. Our computerized version consists of exactly the same items as the handout version and was also split into three parts, the only difference being that the computerized version contained more practice trials so participants could get used to using the keyboard for the procedure. The first part consists of 50 practice trials and 100 experimental trials. The words *Blue*, *Green*, *Red*, and *Yellow* were shown one at a time in black ink on the screen, and participants were instructed to indicate as quickly as possible which word was on the screen using the keyboard. The second part consisted of 25 practice trials and 100 experimental trials. Four pictures of rectangles colored blue, green, red, or yellow were shown on the screen, and participants were instructed to indicate the color of each rectangle as quickly as possible using the keyboard. The pictures remained on the screen until an answer was given. The third part also consisted of 25 practice trials and 100 experimental trials. The words *Blue*, *Green*, *Red,* and *Yellow* were shown in blue, green, red, or yellow ink, with the color word never appearing in the color of ink that the word signified. For example, the word *Red* was shown in green ink. Participants were instructed to indicate the color of the ink (green) as quickly as possible using the keyboard. An interference score was calculated for each participant by abstracting their reaction time of the second part (color naming) from their reaction time of the third part (naming the color of the words). A high interference score indicates low inhibition (Hammes, 1978).

The Swom (spatial working memory) task was designed to measure spatial working memory using Adobe Macromedia Flash Player 8.0. We used a self-designed task, because studies on the role of working memory in planning performance use different working memory tasks and none seemed appropriate to measure the spatial element of working memory that is also required in the TOL. We calculated the internal consistency of the Swom and concluded that it is sufficient (Cronbach's α = .74). The Swom consisted of one example, two practice trials, and 10 experimental trials. All trials showed a grid of nine blocks and 24 paths surrounding the blocks. The items contained two to six series of two blue lines on the paths. All series appeared for 2.7 s and then disappeared. The first series appeared and disappeared, then the second series appeared and disappeared, and finally the last series appeared and disappeared. This appearing and disappearing of the lines resulted in the creation of a movement that had to be kept in memory. The cursor (shaped like a pencil) appeared on the grid with empty paths, and the participant had to reproduce the line series just presented by clicking a path. They were given no instructions about the sequence of their response. The example given to participants is an item consisting of two series of lines. The participant is shown how the series appear and disappear and how the response has to be given using the cursor. The practice trials contained two and four series of lines. The experimental trials contained two to six series of lines; two items with two series of lines, two items with three series of lines, and so on. The items were presented in an increasing degree of difficulty. For each participant, a total correct score was calculated. Depending on the number of correct series, the score could vary from 0 to 40.

Procedure

The three tasks were administered in a random order in one session of 20–45 min, depending on performance. First, oral instructions were given to turn off cell phones and read instructions carefully. Then, each participant could sit at a computer or laptop. Before starting the three tasks, they were asked to fill in their student ID number, date of birth, gender, educational level, and university or academic year.

The instructions of the TOL were given along with a figure showing start and goal positions. The start position was shown in the middle and the goal position was shown in the left upper corner. Participants were instructed to arrange the balls from the start position into the goal position while taking the three restrictions into account: (a) The balls can only be replaced one by one; (b) they cannot be placed outside the pegs; and (c) a maximum of three balls can be placed on the tallest peg, a maximum of two on the middle peg, and a maximum of one on the shortest peg. Participants were instructed to solve the problems in the minimum number of moves, which were given below the goal position. The task was programmed to stop presenting the current problem and move on to the next problem when a move made it impossible to solve the problem in the minimum number of moves. Participants were made aware of this and were instructed to think well before making a move, because they did not get the opportunity to undo a move. This was intended to encourage participants to rely more heavily on planning capacity (Goel & Grafman, 1995; Lehto et al., 2003). After the instructions were given, two practice trials were presented. During the experimental trials, participants also received feedback, but they were not given the opportunity to undo a move. For the Stroop task, a general instruction was given first, indicating that this task consists of three parts and that responses can be given using the keyboard. For the Swom task, instructions were given along with a picture of the grid. After the example was shown, the two practice trials were presented.

Statistical Analysis

Research Question 1: Factor analysis. To investigate whether the variance in correct and incorrect scores on the 12 TOL problems could be explained by the two hypothesized factors (inhibition and spatial working memory), we performed a dichotomous factor analysis because the responses to the problems were measured on a nominal (correct or incorrect) scale. The program Latent Gold (Vermunt & Magidson, 2005) was used to estimate the parameters and calculate the fit of a one- to three-factor model.[1] The likelihood square (L^2) expresses the fit of the model. The amount of reduction of the L^2 for models with an increasing number of factors can be considered in order to choose the best-fitting model, taking the number of parameters into account. A chi-square test on the difference in L^2 can be performed to test the difference between two nested factor models. Moreover, the Bayesian information criterion (BIC; Raftery, 1995; Schwarz, 1978) index, defined as $-2 \times L^2$ + number of parameters $\times$ $\ln(N)$, was calculated for each factor model. The BIC weights the fit and the parsimony of a model. Given any two estimated models, the model with the lower value of BIC is the one to be preferred.

Research Question 2: Latent class regression analysis. A latent class regression model (Vermunt & Magidson, 2005; Wedel & DeSarbo, 1994) was used to investigate whether the strategy groups that differ with respect to the relationship between problem properties of the items and preplanning time on the 12 TOL problems that were administered to all participants could be distinguished.

The latent class regression model can be formulated as follows. Let y_{ik} be the realized value of person i ($i = 1, \ldots, N$) on TOL problem number k ($k = 1, \ldots, K$). The number of problems (within-subject levels) is denoted by K (note that $K = 12$). Let X (realizations x, $x = 1, \ldots, T$) be a latent categorical variable representing the strategy groups. Let z_k be a vector containing categorical predictor variable j representing the problem properties (note that $J = 3$). Then, the first part of the latent class regression model is defined as the probability P of being in a particular latent class x, that is, $P(x)$. These marginal probabilities of being in a specific class add to 1 over the latent classes x. The probability $P(x)$ is also called the class size because it indicates the proportion of the sample that belongs to class x. The second part of the model is a density function for the dependent variable, given the latent class and given the level of the predictor variable: $f(y_{ik}|x,z_k)$. Then, equations $P(x)$ and $f(y_{ik}|x,z_{ik})$ combine into the latent class regression model. The model is defined by a summation over latent classes of products of the marginal probabilities of being in a latent class and the product of class-specific densities for the levels of the predictor variable:

$$f(y_i \mid z) = \sum_{x=1}^{T} P(x) \prod_{k=1}^{K} f(y_{ik} \mid x, z_{ik}). \quad (1)$$

To calculate the multinomial probabilities of being in a latent class, $P(x)$, one parameter, η_x, has to be estimated for each latent class. Note that these η_x parameters sum to zero over latent classes. $P(x)$ is defined by a logistic regression function:

$$P(x) = \frac{\exp(\eta_x)}{\sum_{x=1}^{T} \exp(\eta_x)}. \quad (2)$$

To calculate the second part of the model $f(y_{ik}|x,z_{ik})$, a linear predictor, ν_{x,z_k}, has to be calculated that contains two kinds of parameters, denoted by β_{x0} and β_{xj}, $\nu_{x,z_k} = \beta_{x0} + \sum_{j=1}^{j=J} \beta_{xj} \cdot z_{jk}$. Parameter β_{x0} is the class-specific intercept and parameter β_{xj} is the class-specific regression coefficient. Because the predictor is a nominal variable, there is a parameter β_{xj} for all levels of the predictor variable in each latent class.

Again, the program Latent Gold 4.5 (Vermunt & Magidson, 2008) was used to estimate the parameters and calculate the fit of the model. Second, bootstrap p values were estimated for the -2 log likelihood (LL) difference test between the one- to four-class models. Third, the BIC, defined as $-2 \times$ LL + number of parameters $\times$ $\ln(N)$ was calculated for each model. The BIC weights the fit and the parsimony of a model. Fourth, the proportions of classification errors are provided. This proportion indicates how well the method can predict latent class membership

[1] See http://www.statisticalinnovations.com/products/latentgold_v4.html for more information and a free demo version of Latent Gold.

given the dependent variable (Andrews & Currim, 2003). This proportion is not a fit measure but it is an important measure to evaluate the distinctiveness of different classes. Fifth, the proportion of explained variance is provided.

Results

Research Question 1: Fit of the Factor Model

Table 3 shows the fit of the one- to three-factor models. A chi-square test on the difference in L^2 between the one- and two-factor models was not significant. This means that the second factor did not explain a significant part of the variance in TOL total correct scores. The BIC index was lowest for the one-factor model indicating that this model has the best fit when the number of parameters is taken into account. The correlation between the factor scores of the one-factor model and the score on the Stroop was not significant ($r = .01$, $p > .05$), indicating that inhibition as measured by the Stroop was not related to the factor. The correlation between the factor and the Swom was significant ($r = .229$, $p < .01$), indicating that a small part of the variance in planning performance could be explained by spatial working memory ability as measured by the Swom. The factor loadings of the 12 TOL problems were low (all $< .16$), indicating that most of the variance in planning performance could not be explained by one common factor.

Research Question 2: Fit of Latent Class Regression Model

Latent class regression models containing one to four classes were fitted with preplanning time as the dependent variable and the three problem properties (number of possible initial moves, number of moves, alternative goals) as within-subjects factors. We used preplanning time as dependent variable instead of a correct score, because we instructed participants to solve the problems in the minimum number of moves. Therefore, the preplanning time is bound to provide more useful information about planning than a correct score.

Table 4 shows the fit statistics for the one- to four-class regression models. The log-likelihood increased from the one-class to the two-class model and from the two-class model to the three-class model. This result was confirmed by the bootstraps, which showed significant differences between the one-class model and the two-class model and between the two-class model and the three-class model and no significant differences between the three-class model and the four-class model. These results indicate that the three-class model fitted best. However, the BIC was lowest for the four-class model and the proportion explained variance was 8% higher for the four-class model than for the three-class model. These results indicate that the four-class model fitted best. However, the number of classification errors was three times higher for the four-class model than for the three-class model. To choose a good-fitting model, it can be useful to take a closer look at the interpretation of the models. Therefore, we compared the three-class model and the four-class models in terms of class membership interpretation. The estimated class sizes for the three-class model were .56, .31, and .14 with mean preplanning times of 27 s, 10 s, and 70 s, respectively. The four-class model had class sizes .43, .27, .25, and .05, with mean preplanning times of 23 s, 10 s, 41 s, and 110 s, respectively. Thus, in comparison with the three-class model, the four-class model contained a small additional class characterized by a very long preplanning time. Because this small class seemed to be a typical class that clearly deviates from the other classes, we decided to interpret the four-class model.

Table 5 shows the estimated parameters β_{x0} and β_{xj} for the four classes and the accompanying standard errors. The regression weights of the problem properties had the same direction in all classes. That is, the preplanning time increased with an increasing number of moves and with an increasing number of possible initial moves. However, the preplanning time was shorter for problems with an alternative goal move. The effects of all predictors were strongest in Class 2. In this class, how long participants do their preplanning, which is on average very long, depends strongly on the problem properties. Strong effects for the three predictors were also found for the fourth class, which was also characterized by long preplanning time on average. The effects of the predictors were small for Class 1, which was characterized by a short average preplanning time. Apparently, the short preplanning time of participants in this class was independent of the problem properties.

To investigate class differences with respect to correct performance on the TOL problems, inhibition performance measured by the Stroop, and spatial working memory performance measured by the Swom, we performed a multivariate analysis of variance. The means and standard errors for the three dependent variables are shown in Table 6. The multivariate effect was significant, $F(9,443.1) = 11.25$, $p < .001$. The univariate effects showed that the multivariate effect could be attributed to differences in performance on the TOL problems, $F(3,184) = 35.91$, $p < .001$; the univariate effects for inhibition and spatial memory were not significant, Fs < 1. Post hoc tests (Bonferroni adjustment) showed significant differences (ps $< .05$) in average performance between all groups except for the difference between Classes 3 and 4 ($p > .05$). The mean average TOL score was largest for Class 2, followed by Classes 3 and 4. Performance in Class 1 was the lowest.

Table 3
Fit Statistics for the One- to Three-Factor Models

Model	Npar	L^2	BIC(L^2)
One factor	24	798.57	−60.21
Two factors	36	777.86	−18.08
Three factors	48	760.62	27.51

Note. Npar = number of parameters; L^2 = likelihood square; BIC = Bayesian information criterion.

Discussion

In this study, we expected that, in particular, variation in performance in spatial working memory would explain variation in performance on TOL problems. Variation in performance in inhibition was not expected to explain as much variation in planning performance as spatial working memory, because we used a perfect solution approach in which participants had to reach the goal position in a minimum number of moves. The results of the factor analysis on 12 individual TOL problems showed that one under-

Table 4
Fit Statistics for the One- to Four-Latent Class Regression Models

Regression model	LL	BIC (LL)	Npar	CE	R^2
One class	−11,485.39	23,007.47	7	.000	.049
Two classes	−10,229.34	20,537.31	15	.008	.259
Three classes	−9,698.78	19,518.12	23	.009	.326
Four classes	−9,522.58	19,207.66	31	.026	.408

Note. LL = log likelihood; BIC = Bayesian information criterion; Npar = number of parameters; CE = proportion of classification errors.

lying factor could explain a small part of the variation in performance on these problems. However, most of the variation in performance on the TOL problems remained unexplained. There was no significant relationship with the Stroop's interference score, and the variance in this factor could, for a small part, be explained by spatial working memory ability as measured by the Swom. This result may be a consequence of the fact that we did not take into account the influence of the problem properties and the different strategies.

To investigate the influence of problem properties and the existence of different groups in terms of strategy use, we applied a more detailed approach in the second analysis. We distinguished four latent classes and predicted the influence of problem properties on their preplanning time using a latent class regression analysis. Table 7 shows the characteristics of these four estimated classes and those of the three hypothesized groups. As can be seen in the table, two of the three hypothesized groups resemble a class from the latent class regression analysis. The first hypothesized group, described as not adopting an effective strategy, agreed with the estimated class one. Participants in this class showed a short preplanning time and a small effect of the problem properties. Furthermore, they obtained the lowest scores on the TOL. Therefore, the strategy of this class can indeed be described as no effective strategy. Participants in this class start their first move too fast and make a large number of mistakes. Although this might seem to be an impulsive strategy, in our first analysis inhibition was not found to explain variation in TOL scores and in the second analysis inhibition was not a significant predictor of preplanning time. Therefore, we interpret this finding in terms of another ineffective strategy that does not necessarily relate to inhibition, namely, the trial-and-error strategy. In the manual of the Dutch computerized version of the TOL, Kovács (2007) mentioned that it is possible that participants adopt a trial-and-error strategy. He warranted that this subgroup may resemble clinical subgroups but that these people are, in fact, not concerned with planning during the preplanning time and, instead, make their moves on the basis of trial and error. This characterization seems to coincide with our estimated Class 1. The second hypothesized group, described as having efficient planning ability, agreed with the estimated Class 2. Participants in this class were characterized by the longest preplanning time and the strongest effect of the problem properties. They had the highest score on the TOL and, therefore, the strategy of this class can be described as efficient planning. The third hypothesized group, described as having inefficient planning and characterized by a long preplanning time, low performance on the TOL, and a small influence of the problem properties, was not found in any estimated class. A possible explanation is that this combination of characteristics is not likely to occur in a healthy college population when a perfect solution approach is adopted. In the study of Owen et al. (1990), frontal lobe patients with an inefficient strategy showed a long preplanning time but also a long execution time. However, for our participants, it was not convenient to plan online, because the first move had to be correct; otherwise, the presentation of the problem was stopped. Therefore, they were forced to plan during preplanning time. It turned out that the expectation to find the specific combination of a long preplan-

Table 5
Parameter Estimates and Standard Errors of the Intercept and Predictors

	Class 1		Class 2		Class 3		Class 4	
Intercept or predictor	Value	*SE*	Value	*SE*	Value	*SE*	Value	*SE*
Intercept β_{x0}	8.87	0.28	95.21	14.41	20.95	0.66	36.26	1.81
Predictors β_x								
Moves								
Five	−0.62	0.31	−24.49	16.30	−2.99	0.70	−14.45	1.89
Six	−0.32	0.32	−0.89	16.90	−0.35	0.72	−4.50	1.98
Seven	0.94	0.32	25.38	16.38	3.34	0.70	18.95	2.15
Alternative goals								
No	1.36	0.27	26.99	14.39	2.68	0.60	6.87	1.59
Yes	−1.36	0.27	−26.99	14.39	−2.68	0.60	−6.87	1.59
Possible initial moves								
Two	−1.56	0.36	−31.18	18.63	−5.54	0.82	−13.36	2.10
Three	−0.07	0.32	−11.16	16.85	−0.24	0.70	−0.18	1.88
Four	1.63	0.32	42.35	16.89	5.78	0.73	13.53	1.88

Table 6
Mean and Standard Deviation of Average Tower of London (TOL) Score, Spatial Memory, and Inhibition, Per Class

Class	TOL score		Spatial working memory		Inhibition	
	M	*SD*	*M*	*SD*	*M*	*SD*
1	.29	.16	27.43	5.56	21.83	18.24
2	.82	.18	26.71	4.96	22.59	17.66
3	.53	.18	27.66	4.92	23.48	15.70
4	.61	.20	29.00	5.69	23.36	13.85

ning time and a small effect of the problem properties in our sample was not appropriate, because participants with a long preplanning time are very likely to notice the problem properties and, therefore, be influenced by them.

The results also showed two classes that were not hypothesized but have a number of interesting characteristics. Participants in both classes (Classes 3 and 4; see Table 7) were characterized by an average score on the TOL. However, when preplanning time and the problem properties are taken into account, the classes seem to differ with respect to strategy use. More specifically, participants in Class 4 think longer before making their first move than do participants in Class 3, but they do not perform significantly better in terms of correct score. Their mean TOL score may be somewhat higher (see Table 6), but it is not high enough to be interpreted as having better planning ability than participants in Class 3. Also, in Class 4, the preplanning time is strongly affected by the problem properties, whereas in Class 3, the effect of the problem properties is average. It seems that participants in Class 3 plan their moves more efficiently during the preplanning time and are less distracted by the problem properties than are participants in Class 4. This finding illustrates the importance of calculating more than just a total correct score and the influence of problem properties on different subpopulations in a sample.

The results of this study also showed that in all the classes, preplanning time increased with an increasing number of moves and with an increasing number of possible initial moves and that the preplanning time is shorter for problems with an alternative goal move. Except for the two average-scoring classes, the estimated classes differed in performance on TOL scores. However, the four classes did not differ in performance on Stroop and Swom scores and performance on neither task was a significant predictor of preplanning time. The effect sizes of the nonsignificant differences were too small to be interpreted as tendencies.

There are at least two explanations for the finding that inhibition was not related to TOL performance and preplanning time. First, participants did not get the opportunity to undo a move. In our instructions, we asked for solutions within the minimum number of moves to prompt participants to plan their moves carefully (Berg & Byrd, 2002; Goel & Grafman, 1995). In fact, participants were instructed to inhibit and, therefore, this approach may not differentiate between participants' inhibition ability. Moreover, in the study of Huizinga et al. (2006), performance on the Stroop was also not related to preplanning time on the TOL. Second, it is questionable whether the Stroop task measures the same type of inhibition as the TOL is assumed to measure. The interference score of the Stroop indicates the speed of inhibition, whereas in the TOL it is important to inhibit an incorrect move and the speed of inhibition is not an issue. Therefore, the interference score of the Stroop and the preplanning time on the TOL may operate differently on the construct of inhibition.

Although we did expect that performance on the Swom would play a greater role in TOL performance than the results showed, we have a suggestion for why it was not a predictor of preplanning time in the latent class regression analysis. Spatial working memory as measured by the Swom seems to tap especially the spatial aspect of the recalling of the plan constructed during preplanning time. However, the spatial recollection during preplanning time concerns only the recollection of the first move. Therefore, it may be expected that spatial working memory would have a greater influence on the execution time than on the preplanning time. If this hypothesis can be confirmed, it would lead to a clearer description of the cognitive processes involved in TOL perfor-

Table 7
Summary of the Hypothesized Groups and the Estimated Classes

Group	Class size	Preplanning time	Performance	Influence properties	Inhibition	Spatial working memory
Hypothesized						
No effective strategy		Short	Low	Small	Average	Unknown
Efficient		Long	High	Large	Average	High
Inefficient		Long	Low	Small	Average	Low
Estimated						
No effective strategy	.27	Short	Low	Small	Average	Average
Efficient	.05	Long	High	Large	Average	Average
Average efficient	.43	Average	Average	Average	Average	Average
Average inefficient	.25	Long	Average	Large	Average	Average

mance: (a) When a perfect solution approach is applied, inhibition does not seem to play a role in the actual planning; (b) the actual planning of the moves therefore takes place during preplanning time; and (c) during the execution of that plan, working memory plays an important role, because the plan has to be held in memory while the balls are being moved. This issue should be addressed in future research. We suggest this could be investigated using an approach similar to the one used in this study. For example, by calculating a preplanning time, an execution time, and a correct score per TOL problem and using multiple external measures of spatial working memory, one could investigate the relationship between spatial working memory performance on the one hand and the different scores per TOL problem on the other hand.

Conclusion

The in-depth analysis on TOL-problem level as described in this study confirms just how complex planning as a construct and the TOL as an instrument actually are. It seems that a low correct score on the TOL does not simply mean that one has low planning ability. Moreover, it is unclear what "low planning ability" means in terms of cognitive capacities. Multiple factors influence performance on the TOL and the interpretation of a score. First, the way in which the TOL is administered is related to different cognitive processes. For example, inhibition is related to TOL performance when participants are allowed to make more than the minimum number of moves, but when one uses a perfect solution approach as in this study, inhibition does not play a significant role. Also, the amount of obtained information about a person's planning ability depends on the calculated scores.

In this study, we showed that several scores per item give more information than a total or average score. First, the combination of a correct score and a (preplanning) time score portrays the efficiency of certain planning strategies. Second, the presented TOL problems have different properties, which influence performance. For example, the presence of an alternative goal move is related to a shorter preplanning time. Intuitively, this would lead to a greater probability of obtaining an incorrect score, because the problem is not thought through. Third, people adopt different planning strategies that may depend on these problem properties and also influence performance. These strategies provide useful information about the cognitive processes involved in planning. As the results of this study have shown, people can have roughly the same correct score, but, by disentangling the different factors that influence strategy use, we found that some average-scoring people can still show more efficient planning than others by solving the TOL problems in a shorter preplanning time. In clinical practice, all of the scores are used to describe a patient's strengths and weaknesses when solving TOL problems. This study offers empirical support for such an administration of the TOL that should also be adopted in scientific research. However, it has to be clear, then, which TOL problems should be chosen and which scores should be calculated to measure specific cognitive processes, given certain properties of the problems and the strategies of individuals presented with the TOL. Therefore, it is important to know exactly what the interpretation of certain scores is to draw valid conclusions about performance.

A limitation of this study is that the sample was rather homogeneous and the scores of spatial working memory and planning ability showed little variance. Therefore, our results may not be completely generalizable to other subgroups. It seems quite useful to investigate to what extent the present results apply to other subpopulations, like children or a clinical sample, because it could provide valuable information about learning styles or it could help in the diagnosis of, for instance, attention-deficit/hyperactivity disorder. Future research could study whether strategies used by various groups on the TOL are predictive of strategies used in everyday life, which would help researchers understand the complexity behind planning ability and its applications.

To summarize, on the basis of the results of this study, we offer the following recommendations about the administration of the TOL to enhance the validity of score interpretations. First, a total score, as the total move score or total correct score, does not seem to be a valid measure to capture planning ability and should therefore not be used as the primary score to measure performance, especially when investigating the way people solve problems as in the TOL. Therefore, we recommend using a combination of the preplanning time and a correct score per item. By using these scores, one can differentiate between the subgroups that exist in a sample, which can be characterized by a strategy that may be the result of efficient or inefficient planning. Second, when using the TOL, the influence of the problem properties should be taken into account, because the way in which performance and strategy use are affected by them gives useful information about planning ability. Finally, more research is needed to identify the strategies people adopt when confronted with the TOL, just as strategies for the TOH have been identified (Goel & Grafman, 1995). We postulate that future researchers should also adopt an individual differences approach while taking into account the different problem properties, for such a detailed analysis is crucial when investigating a construct as complex as planning.

References

Andrews, R. L., & Currim, I. S. (2003). A comparison of segment retention criteria for finite mixture logit models. *Journal of Marketing Research, 40,* 235–243.

Berg, W. K., & Byrd, D. L. (2002). The Tower of London spatial problem-solving task: Enhancing clinical and research implementation. *Journal of Clinical and Experimental Neuropsychology, 24,* 586–604.

Culbertson, W. C., & Zillmer, E. A. (2001). *Tower of London—Drexel University.* North Tonawanda, NY: Multi-Health Systems.

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

Hammes, J. G. W. (1978). *De Stroop Kleur-Woord Test: Handleiding* [The Stroop color–word test: Handbook]. Amsterdam, the Netherlands: Harcourt.

Hodgson, T. L., Bajwa, A., Owen, A. M., & Kennard, C. (2000). The strategic control of gaze direction in the Tower of London task. *Journal of Cognitive Neuroscience, 12,* 894–907.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44,* 2017–2036.

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

Kaller, C. P., Unterrainer, J. M., Rahm, B., & Halsband, U. (2004). The impact of problem structure on planning: Insights from the Tower of London task. *Cognitive Brain Research, 20,* 462–472.

Kovács, F. (2007). *Tower of London test voor Windows 9X/ME/2000/NT/XP versie 3.0: Handleiding* [Tower of London test for Windows 9X/ME/2000/NT/XP Version 3.0: Handbook]. Voorhout, the Netherlands: Pyramid Productions.

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21,* 59–80.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.

Murji, S., & DeLuca, J. W. (1998). Preliminary validity of the Cognitive Function Checklist: Prediction of Tower of London performance. *The Clinical Neuropsychologist, 12,* 358–364.

Owen, A. M. (1997). Cognitive planning in humans: Neuropsychological, neuroanatomical and neuropharmacological perspectives. *Progress in Neurobiology, 53,* 431–450.

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesion in man. *Neuropsychologia, 28,* 1021–1034.

Poole, D. G. (1994). The towers and triangles of Professor Claus (or, Pascal knows Hanoi). *Mathematics Magazine, 67,* 323–344.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111–163.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 298,* 199–209.

Vermunt, J. K., & Magidson, J. (2005). *Latent Gold 4.0 user's guide.* Belmont, MA: Statistical Innovations.

Vermunt, J. K., & Magidson, J. (2008). *Latent Gold 4.0* [Computer software]. Belmont, MA: Statistical Innovations.

Ward, G., & Allport, A. (1997). Planning and problem solving using the five disc Tower of London Task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 50*(A), 49–78.

Wedel, M., & DeSarbo, W. A. (1994). A review of recent developments in latent class regression models. In R. P. Ragozzi (Ed.), *Advanced methods of marketing research* (pp. 352–388). Cambridge, MA: Blackwell.

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.

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.

Received September 17, 2009
Revision received February 19, 2010
Accepted June 10, 2010 ■