The Tower of London Spatial Problem-Solving Task: Enhancing Clinical and Research Implementation

Article *in* Journal of Clinical and Experimental Neuropsychology · September 2002

Impact Factor: 2.08 · DOI: 10.1076/jcen.24.5.586.1006 · Source: PubMed

2 authors:

W. Keith Berg
University of Florida

Dana Byrd
Texas A&M University - Kingsville

Journal of Clinical and Experimental Neuropsychology
2002, Vol. 24, No. 5, pp. 586–604
1380-3395/02/2405-586$16.00

The Tower of London Spatial Problem-Solving Task: Enhancing Clinical and Research Implementation

W. Keith Berg and Dana L. Byrd
Department of Psychology, University of Florida, Gainesville, FL, USA

ABSTRACT

Since its development in 1982, The Tower of London (TOL; Shallice, 1982) spatial problem-solving task has been increasingly employed in test batteries of executive functions. This task has served as a rich source of information on preparation, planning and processing, but a number of issues remain unaddressed in the literature: (1) the problem structure, or problem space of the task, (2) the impact of modifications from the original, Shallice TOL, and (3) the variety of performance measures that can be derived from the TOL. We present here an overview of these issues in the hope that it may lead to a more effective and reasoned use of the TOL task by clinical and nonclinical investigators, alike.

In his 1982 study of the planning deficits displayed by frontal lobe lesion patients, Shallice (1982) proposed and used an alternative to the classic Tower of Hanoi (TOH) task. As a native of England, Shallice dubbed the proposed replacement to the TOH, the Tower of London (TOL). The TOL task, now frequently used to study planning ability in both clinical (e.g., Cockburn, 1995; Levin et al., 1996; Murji & DeLuca, 1998) and nonclinical (e.g., Dagher, Owen, Boecker, & Brooks, 1999; Kafer & Hunter, 1997; Morris, Ahmed, Syed, & Toone, 1993) populations, shares with the TOH the goal of rearranging a small set of distinct objects arranged on three rods from their initial position into another specified configuration (see Fig. 1 for illustrations of the TOH and Shallice's TOL). To solve the TOH, disks of varying diameters are moved one at a time from one rod to another with the restriction that no larger disk may be placed on a smaller disk. To solve the TOL, a set of three balls or beads differing in color are moved one at a time from one rod to another with the restriction that each of the three pegs of descending lengths could hold only 3, 2, or 1 balls, respectively.

Shallice said little of why he preferred his own task to that of the TOH other than that it allowed him "to produce a graded difficulty test" (Shallice, 1982, p. 204) and that the TOL allows for a greater variety of qualitatively different problems (Shallice & Burgess, 1991). Importantly, Shallice presented evidence of TOL performance deficits specific to left anterior prefrontal lesions. This evidence and results from several studies that followed (e.g., Morris et al., 1988; Owen, Downes, Sahakian, Polkey, & Robbins, 1990) have demonstrated the value of the TOL in assessing problem-solving generally, and planning specifically, with a wide age range of patients and nonclinical controls (e.g., Anderson, Anderson, & Lajoie, 1996; Cockburn, 1995; Hughes, 1998; Lange et al., 1992). The result is that the TOL has become a very popular spatial problem-solving task in studies of frontal lobe dysfunction in a variety of clinical populations (e.g., Lange et al., 1992; Owen et al., 1990, 1995),

Address correspondence to: W. Keith Berg, SF, Psychology Building, Department of Psychology, College of Liberal Arts and Sciences, P.O. Box 112250, Gainesville, FL 32611-2250, USA. Tel.: +1-904-392-6608. E-mail: Berg@psych.ufl.edu
Accepted for publication: January 25, 2002.

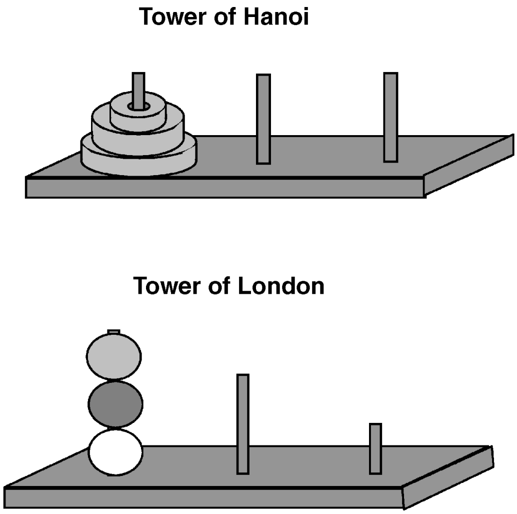

Fig. 1. Illustrations of the Tower of Hanoi and Tower of London (Shallice, 1982). The Tower of Hanoi disks typically differ in color. The goal of the Tower of Hanoi is to move the tower of disks to a different peg or to match a pictured disk arrangement (not shown here). Participants are instructed that a larger disk must not be placed on a smaller one. The Tower of London balls are typically green, blue, and red. The goal of the Tower of London is to match the pattern and color of balls to a pictured arrangement (not shown). In doing this, no more than two balls may be placed on the middle peg, and no more than one ball may be placed on the smallest peg.

and versions of this task are increasingly appearing in standardized batteries of executive function (e.g., Cambridge Neuropsychological Test Automated Battery, Colorado Neuropsychology Tests and NEPSY).

Despite the recent popularity of the TOL task, there has been a notable lack of uniformity in procedure and measurement across both cognitive and clinical studies, although there have been some attempts at standardization (e.g., Anderson et al., 1996; Culbertson & Zillmer, 1998b, 1998c; Schnirman, Welsh, & Retzlaff, 1998). When reviewing the literature on the TOL task, one finds a very wide variety of problem selections, performance measures, and test methodologies, with little or no rationale provided as to the choice of these parameters. This inconsistency could come from a variety of sources: (1) lack of information about the basic TOL structure, (2) little focus on the determinants of problem difficulty, and (3) insufficient information on the possible performance measures. Without such information, there is very little on which to base one's specific selection of problems or performance measures. This leaves each researcher to either make an ad hoc decision on evaluation parameters or to duplicate the selection of a previous study in hopes of gaining some basis for comparison. Additionally, a number of variations on the original version of the TOL have been developed, and these variations may serve as an additional source of confusion. The consequence of this lack of consistency and rationale is that it is very difficult, if not impossible, to directly compare results across studies, to generalize results, or to effectively evaluate apparently contradictory results. The intent of this paper is to provide an overview of these issues and recommend some solutions.

THE IMPORTANCE OF THE PROBLEM SPACE

A logical starting point to finding a rationale for the selection of problems, procedures, and scoring is the underlying problem structure of the TOL task. The graphic representation of the moves possible under the rules of the task is called the task's "problem space." The problem space provides a diagram of every possible legal position of the game pieces connected by lines that represent the legal moves that can be made to move from one to another of these positions. Essentially, the problem space is a road map indicating how to get from any legal position to any other legal position through legal moves.

An understanding of the problem space allows one to plot the optimal path or paths for solving any possible problem and to visualize the potential incorrect moves that could be made. If one also has the timing of a participant's every problem-solving move, then one can ascertain at which points during the task solution there was more extended consideration of the next move (i.e., planning or problem analysis).

This information, both about problem structure and solution paths, could have important clinical and cognitive research uses. One could find that some deficits in specific aspects of problem solving are more readily discernable from the participant's movement through the problem space. For instance, one might note that pauses consistently occur when the participants are faced with specific configurations of the balls. This would suggest that such configurations present a unique cognitive challenge to the participants. If so, a clinician or researcher could select a set of problems from the problem space that would assess this specific challenge and compare this with a control set of other problems without this challenge. This process could be done without the problem space, but would be considerably more difficult.

Problems whose optimal solution (solution in the least number of moves possible) requires removal of an obstacle (i.e., one ball has to be moved out of the way to rearrange the other balls) provide this sort of challenge to participants. Selection of such problems and the contrasting problems that do not have obstacles is greatly aided by the use of the problem space. This becomes especially valuable, for example, when one wishes to make this contrast while holding the number of moves constant. By holding the number of moves constant, the researcher is able to keep working memory requirements (the number of moves) equal between the two sets of problems, while manipulating the requirement of means–end problem solving (problems with and without an obstacle). This type of problem manipulation has proven useful in the examination of children's problem solving with both the TOH (Klahr, 1985) and TOL (MacDonald, Garner, & Spurgeon, in press). Evidence from our laboratory and others suggests that problem manipulations like these also impact normal adult solutions to such problems (Berg & Byrd, in preparation; Ward & Allport, 1997) as well as differentiate between mentally retarded and non-mentally retarded groups' solutions (Spitz, Minsky, & Bessellieu, 1984).

Another illustration of the value of the problem space in problem selection is the ability to select problems with particular ending and starting arrangements. Problems differing in this manner require a different ordering of subgoals. For example, one could compare problems with flat starting positions, one ball on each of the three pegs, to problems with tower starting positions, all three balls on the tall peg. Problems with tower starting positions have a more obvious initial subgoal ordering than problems with flat starting positions since the former have only two possible initial moves and the latter four possible initial moves. Also, selecting problems with a tower *end* position clearly orders the last moves for the position, as compared to flat end positions that require an unclear final sequence of moves.

The structure of the problem space is determined by the configuration of the game elements and its rule structure. With the TOL, a listing of all possible positions with three differently colored balls yields 36 unique positions. The 36 positions can be broken down into a 6 by 6 matrix of six spatial arrangements of balls by 6 different color permutations of these six positions (see Table 1 for an illustration of all six spatial arrangements and two color permutations of these). In the 6×6 table, each column of color permutations is one side of the problem space, making these 36 positions effectively represented with a hexagonal shape[1] (see Fig. 2 for the TOL problem space diagram). In this hexagonal problem space, each of the six faces of the problem space contains the six legal positions for one color permutation.

An understanding of this organization provides a number of valuable practical benefits, as well as revealing some important discoveries about the TOL problem set. At the practical level, the organization suggests a logical numbering scheme for describing problems, ball positions, and moves. In Figure 2, each of the six faces of the hexagon is given a number, 1 through 6, corresponding to one of the six color permutations. Within each face are the six positions of the balls that have also been labeled 1 through 6 (see Table 1). Positions with corresponding spatial arrangements are given the same digit on all six faces (permutations) of the problem space (e.g., the tower position – all balls on longest peg – is

[1]As a comparison, the TOH problem has a triangular problem space that, with a three-disk tower, includes 27 positions.

Table 1. A Representation of the TOL Organization. Shown are the Six Spatial Arrangements of the Balls and Two of the Six Color Permutations that Underlie the Six-sided Tower of London Problem Space. The Six Possible Ball Arrangements, Ignoring Ball Colors, are in the Left Two Columns along with the Appropriate Arrangement Numbers. In the ‘‘Arrangement’’ Column, Balls are Indicated by an ‘‘O,’’ and are Shown as if on Pegs. An ‘‘–’’ Indicates an Open Space on the Peg. Shown in the ‘‘Permutation’’ Columns are Ball Colors for Two of the Six Possible Permutations in the Problem Space. In these Columns the letters ‘‘B,’’ ‘‘G,’’ and ‘‘R’’ Refer to Blue, Green, and Red balls, Respectively. Each Position in the ‘‘Permutation’’ Columns Corresponds to an Arrangement in Left Column. The Particular Permutations Shown also Correspond to One of the Hexagonal Sections of the Problem Space: Permutation Column 1 Corresponds to the Right, Middle Section of the Problem Space and Permutation Column 2 Corresponds to the Bottom, Right Section of the Problem Space (See Fig. 2). Note that each Position in Permutation 2 Differs from Corresponding Positions in Permutation 1 Only in that the Blue and Green Balls are Exchanged. The Remaining Four Color Permutations (Not Shown) Correspond to One of the Four other Sections of the Hexagonal Problem Space (See Fig. 2). These are Achieved by Simply Consistently Reordering the Colors in Permutations 1 and 2.

Spatial arrangements		Permutation ‘‘1’’		Permutation ‘‘2’’	
Arrangement	Number	Ball colors	Position #	Ball colors	Position #
O	01	B	11	G	21
O –		G –		B –	
O –		R – –		R –	
–	02	–	12	–	22
O –		G –		B –	
O – O		R – B		R – G	
–	03	–	13	–	23
O –		G –		B –	
O O –		R B –		R G –	
–	04	–	14	–	24
– O		– G		– B	
O O –		R B –		R G –	
–	05	–	15	–	25
–		–		–	
O O O		R B G		R G B	
–	06	–	16	–	26
– O		– R		– R	
– O O		– B G		– G B	

designated position 1 on all faces). We can uniquely designate any one of the positions with a two digit number by placing the color permutation (face) number in the 10s place and the position number in the 1s place (see numbers on the TOL problem space, Fig. 2 and also Table 1). For instance, the ‘‘tower’’ positions are labeled 11, 21, 31, 41, 51, or 61, depending upon the permutation of ball colors. Similarly, the ‘‘flat’’ positions (one ball on each peg) are numbered 15, 25, 35, 45, 55, or 65.

With such a numbering system one can also efficiently identify any problem presented to the participant by indicating the start and goal position numbers, separated by a hyphen. So, if we wished to designate a problem that started with the tower position of permutation 1 face and had a goal of a ‘‘flat’’ position on the permutation 3 face, the problem would be designated an 11-35. Similarly, we can designate any particular move made by a participant by indicating the numbers of initial position and the new position, separating the pair members with a colon, say, rather than a hyphen to designate a single move. To then indicate that a participant was presented with a problem with the start position 31 (a tower) and the goal position 34 (a two-move problem), and solved it with the optimal (minimum move) path would produce moves 31:33 and 33:34. However, if the participant instead solved this problem

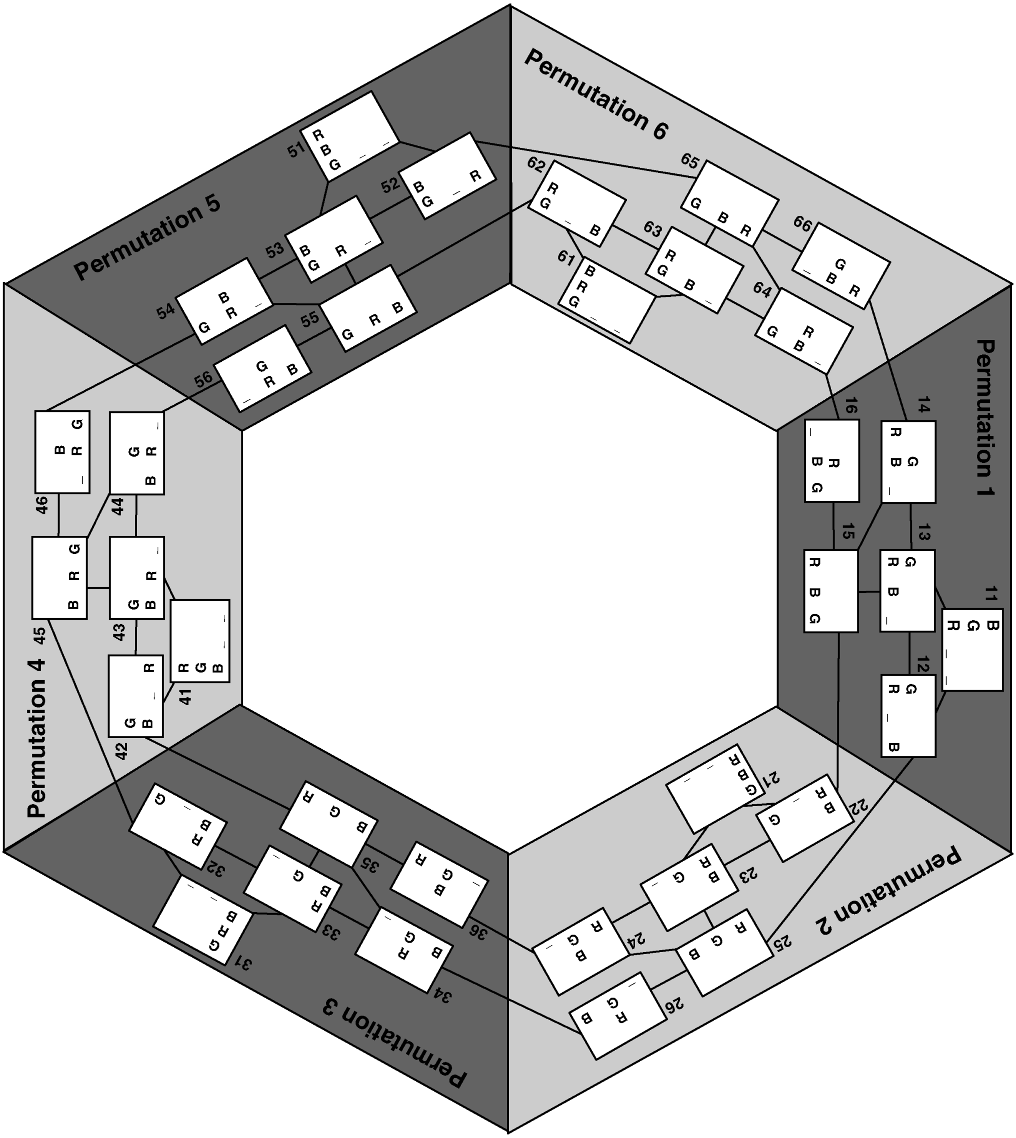

Fig. 2. The hexagonal problem space for the Tower of London and the proposed numbering scheme. The letters B, R and G refer to blue, red and green balls, respectively. Note that the patterns on each face are identical except that the arrangements of the ball colors is permuted and that adjacent faces have the sets of patterns rotated. Thus, for example, the tower position (all three balls on the left peg – positions 11, 21, etc.) occur are at the top of three faces and the bottom of the others.

nonoptimally in three moves, the following solution could occur: Problem 31-34, Moves 31:33, 33:35, 35:34. By examining the problem space figure we can see that the one extra move in this less efficient solution was 33:35.

Using this numbering system, Table 2 shows the original set of problems employed by Shallice (1982), which was also employed by a number of other researchers (e.g., Anderson et al., 1996; Krikorian, Bartok, & Gay, 1994; Owen et al., 1990). The clarity and succinctness of this numbering system will allow investigators, clinicians, and test developers to conveniently specify TOL problems, and will encourage more

Table 2. The 12 Problems Employed by Shallice (1982) using the TOL Problem Space Numbering System Described in the Text and Illustrated in Figure 2.

Starting position for all problems:

Configuration*	Position number**
R G B –	63

Goal Positions for the 12 problems:

Problem sequence***	Configuration*	Position number**	Minimum moves
1	G – B R	66	2
2	G R B B	55	2
3	G R – B	53	3
4	G R – G	54	3
5	R B – G	13	4
6	B R – B	44	4
7	R – G B	22	4
8	– R G B G	46	4
9	R – – G B	11	5
10	R – – G	21	5
11	B R –	43	5
12	B R G	45	5

Note. *Configuration refers to the position of the balls on the pegs, where "B" = blue ball, "G" = green ball, "R" = red ball, and "–" = empty peg. The tallest peg is to the left.
** Position number refers to the number based on the problem space figure in Figure 2.
*** The problem sequence is the same as used in Shallice (1982).

complete reporting of this critical information in academic publications.

The availability of the problem space has value beyond that of providing a succinct numbering scheme. Each TOL problem is one of a set of six problems, all of which require an identical sequence of moves, and differ only in the color of the balls moved. For example, using the problem space figure (Fig. 2), one can see that the optimal solution path for problem 11-25 requires the identical sequence of moves as does the optimal solution path for problems 31-45, 51-65, etc.[2] That is, in all cases one moves the top ball of the tower to the small peg, and then the next-to-top ball to the middle peg. We shall refer to these color-permuted, parallel-move problems as "iso-problems." These sets of six iso-problems are of considerable value for both clinicians and researchers who wish to present multiple problems while controlling precisely for problem difficulty, the movements required, and the consequent cognitive challenges resulting from these. Our laboratory has found that nonclinical, college participants are not aware that such iso-problems require the same moves when the iso-problems are intermixed into a set of other problems. This is true even when the repetition occurs within a single short testing session. Also, we have found that these color permuted iso-problems were of equal difficulty for these participants (Berg & Byrd, in preparation). However, since this has only been informally examined at this time, future research should carefully investigate the equivalence and awareness of iso-problems.

We shall refer to all problems that are not iso-problems as defined above as "spatially unique problems." For two problems to be considered spatially unique, at least one move of their minimum-move solutions must differ from one another. For example, the two three-move pro-

[2]One might think that the three other iso- (identical) problems in this set might be 21-35, 41-55, and 61-15. Careful inspection of the problem space (see Fig. 2) indicates that one must move in the opposite direction around the hexagon to produce the other three iso-problems. Thus, the other three iso-problems for this example are 61-55, 41-35 and 21-15.

blems 11-26 and 11-23 have similar solutions, differing only in the last move, 25:26 in the former, 25:23 in the latter (see Fig. 2). Because they differ in at least one move, they would be considered spatially unique. One can determine the complete set of spatially unique problems possible in the TOL task by starting from each of the six positions on one of the permutation faces of the problem space and determining the solution to every other possible position on all faces of the problem space. In doing this one finds that there are 210 spatially unique problems, ranging from 1 to 8 moves in the optimal solution path. Because of the structure of the TOL task, 8 moves are the maximum number of moves possible in an optimal solution of any possible problem. The set of 210 spatially unique problems is roughly equally distributed for problems of 2 to 7 minimum moves, with a range of 28 to 35 problems in each category. There are fewer 1- and 8-move problems, 18 and 11, respectively. Since each of the 210 spatially unique problems has 5 iso-problems that match it exactly in the necessary moves to achieve an optimal path solution, the total number of all possible problems for the TOL is 1260 problems, including both spatially unique and iso-problems. Thus, the TOL task produces a large and complex set of problems, containing both iso-problems and spatially unique problems. This means that studies using this task have the possibility of both control and flexibility during testing.

Another benefit in problem selection that can be gained from knowledge of the problem space is the selection of "nested problems." One problem is nested into another when the optimal path for the first problem is completely contained within the optimal path for the second problem. For example, the problem 11:25 is nested in problem 11:23, the former containing the same first two moves along the optimal solution path as the latter, and the latter adding one additional move. Thus, a problem may be nested into another by adding additional moves before or after the nested problem. Nesting problems allow one to add working memory demands (additional moves) to a problem, while holding a number of other problem characteristics constant. Also, it is possible that presenting nested problems may impact the participants' rate of learning across the session. However, the impact of nesting on learning, to our knowledge, is yet to be assessed.

VARIANTS OF THE TOL: WHAT'S IN A NAME?

Since the TOL was developed by modifying the TOH, it is not surprising that, in this same spirit of innovation, investigators have in turn devised modifications of the TOL. As valuable as these modifications might be, they can produce a good deal of confusion, especially when new names are introduced or old names are used for new procedures. Such confusion of terminology can turn the Tower of London into the Tower of Babel! For our purposes, we shall refer to these modifications of Shallice's TOL as "variants" even though some do not vary much and others introduce major changes. Examples of this range of variants include the TOL^{DX} (Culbertson & Zillmer, 1998a, 1998b, 1998c), the five-disk Tower of London (Ward & Allport, 1997), the 4-rod Tower of London (Kafer & Hunter, 1997), and the TOL–R (Schnirman, et al., 1998; Welsh, Satterlee-Cartmell, & Stine, 1999).

Some of these variants are actually "iso-morphs" of the TOL. An iso-morph of a task is one that has exactly the same problem space but testing is done using an apparatus with a different appearance. For example, a well-known iso-morph of the TOH is the so-called "monsters-and-globes" task (Kotsovsky & Simon, 1985). It is important to avoid confusion between iso-*morphs* and the iso-*problems* described earlier. The latter involve comparable problems using the same apparatus. They do not require any change in the overall appearance of the test apparatus, only the permuting of ball colors of the test problem. Iso-morphs of the TOL include the Cambridge Neuropsychological Test Automated Battery (CANTAB) version of the TOL (Owen et al., 1990).

It is also possible to produce a task using an apparatus with a generally similar appearance but with differences that produce a fundamentally different problem space. In such a case, the task would be neither an iso-morph nor would it have iso-problems corresponding to Shallice's TOL, but would simply be a different task. We should note that all of such tasks likely have their own unique advantages and can provide valuable information in their own right. However, to avoid confusion when comparing and contrasting results from various tasks with similar names, it is imperative to understand when and how such tasks differ from one another.

To begin to try to organize and evaluate these many variants, we will first briefly describe each of them, considering both the physical appearance and the problem space of the task, and compare them particularly with Shallice's (1982) original version of the TOL. Then we will provide suggestions and recommendations for both clinical and research uses of the TOL.

First, let us clarify that the TOL task as developed by Shallice (1982) is the focus of the present paper. To prevent confusion, we propose that the phrase "Shallice TOL" task be used to clearly distinguish Shallice's original TOL task from TOL variants with similar names. TOL tasks with minor variations, such as in the color or size of the balls, the particular problem selected for testing, the measure of performance, or the instructions would still, in our convention, qualify as the Shallice TOL, as would computerized versions of the task that depict the basic Shallice apparatus. Certainly some of these variations could alter performance. However, we propose to reference any task having an identical problem space to the Shallice's original TOL task and the same basic appearance as that task, as a Shallice TOL task. A Shallice TOL task, as we define it, may or may not employ the set of 12 problems used originally by Shallice.

When, then, should a different name be employed for a variant? Altering the appearance of the task in a major way, but not the problem space of the task, qualifies the new task as an iso-morph of the Shallice TOL. Given the evidence that iso-morphs can very markedly alter performance of a spatial task (e.g., Kotsovsky & Simon, 1985), we propose that iso-morphs should either be referenced using a different name, as is done with iso-morphs of the TOH, or that a clear modifier be used along with "TOL." And finally, when the task is altered such that the resulting problem space is different, we propose that a new name be used, just as Shallice proposed the new name, "Tower of London," for his new task to distinguish it from the TOH. Given these considerations for nomenclature, we will review the various variants and classify them according to this scheme.

Culbertson and Zillmer (1998a, 1998b, 1998c) developed a variant of the TOL they refer to as the TOL^{DX}. This variant has a somewhat different set of problems from those used originally by Shallice (1982) and standardized instructions for participants, an important feature. The manual for the commercially available version (Culbertson & Zillmer, 1998c) includes valuable normative and reliability data as well. Nonetheless, the basic rules, appearance and problem space of the TOL^{DX} are like the original Shallice version. Thus, by our definition, the TOL^{DX} still qualifies as an example of the Shallice TOL.

Schnirman et al. (1998) made a more substantial change to the Shallice TOL when they evaluated a variant they called the TOL – revised (TOL–R). Again, the basic rules, appearance and problem space remain unchanged. However, these authors did provide an assessment of a different and much larger set of problems than the original 12 used by Shallice, and then selected a subset from this larger set based on reliability information. This new problem set appeared to have considerable psychometric advantages, with impressive reliability estimates. Since the basic task itself remains unchanged, this variant is an example of the Shallice TOL.

Sometimes tasks differing physically from the Shallice TOL do not use modified names, potentially leading to confusion. A computerized version of the TOL that has been used in numerous studies by Owen et al. (1990, 1995) qualifies as an iso-morph of the Shallice TOL. Rather than creating a computer image closely mimicking the Shallice TOL beads and rods version (see Fig. 1), Owen and colleagues generated what they named the "pockets" version of the TOL (Owen et al., 1990). In this version balls are still present, but instead of being placed on rods standing upright from a base, they are put into objects that appear similar to "socks" or billiard (or snooker) ball pockets that hang downward from an apparent support. There is a clear notch between where each ball can fit into the pockets. This pockets version of the TOL also utilized a very different manner of ball movement than is used with the physical Shallice TOL. Rather than requiring the actual movement of the simulated balls from pocket to pocket with a drop and drag motion, for example, the balls are moved from one pocket to another with a sequence of two "touches" to the touchscreen monitor: the initial touch selects a pocket from which the top most ball will be moved, and a second touch indicates the pocket into which a ball will be moved, into the lowest open position. Thus, the actual movement of the ball is carried out by the computer.

These changes in procedure may have a number of advantages. In this "pockets" version, with its clear ball placement positions, it may be more obvious how many balls can fit in a pocket. Also, the two-touch ball movement procedure has the advantage of minimizing the amount of hand motion needed to move the balls, a particular value with some disabled groups. However, there is a possible downside to these changes as well. Given the evidence of the powerful effects of isomorphs (Kotovsky, Hayes, & Simon, 1985) on the ease of problem solving, caution is appropriate when comparing results from this version to the Shallice TOL versions.

Kafer and Hunter (1997) compared results for the Shallice TOL with a variant, called the "4-rod TOL." In this variant, a fourth ball and fourth rod that holds four balls, is added to the Shallice TOL. The rest of the apparatus remains unchanged as do the basic rules (Kris Kafer, personal communication, September 6, 1999). This is an interesting variant since it provides a greater range of problem difficulty. However, one important change could make it more compatible with the Shallice TOL. In the Shallice TOL the number of open spaces is equal to the number of balls, which is the most restrictive ratio possible in a playable game.[3] Kafer and Hunter's 4-rod variant has more open spaces than balls, six open spaces and four balls, allowing for a larger number of optimal paths to solution. By modifying Kafer and Hunter's 4-rod TOL to maintain the 1:1 ratio of balls to open spaces it is possible that the resulting task may require more careful planning to achieve a solution with an optimal path.

The TOL variant with the largest and most significant change is the task called the "five disk Tower of London," originated by Ward and Allport (1997) and adopted by other researchers (e.g., Phillips, Wynn, Gillhooly, Della Sala, & Logie, 1999; Gillhooly, Phillips, Wynn, Logie, & Della Sala, 1999). For this variant three changes in physical appearance were made to the Shallice TOL task: (1) disks of equal size were used rather than balls, (2) four to five game pieces were used rather than three, and (3) all pegs were of equal height and each was able to hold all of the disks. The change from balls to disks is a minor physical appearance change, and leaves the basic task unaffected. But the latter two changes result in an entirely different problem space.

The ability to place all five disks on any of the three pegs makes this task somewhat like the TOH; however, using a color-matching goal makes this task somewhat like the TOL. Thus, the new task is something of a hybrid between these two tower tasks. Unlike either the TOH or the Shallice TOL, there is no restriction on the rearrangement of disks either by space limitations or by a size placement limitation. As noted above, the movement restriction for TOL stems from how many balls each peg can hold, which results in the number of open spaces exactly equaling the number of balls. The restriction on disk movements in the TOH results from the rule that a larger disk cannot be placed on a smaller one. The removal of both such restrictions not only changes the problem space but may also change the cognitive requirements of the task as the participant does not have to remember that rule.

These changes in the five disk task, both in structure and in rules, result in a unique task with its own beneficial qualities, as is evident from the studies by Ward and Allport (1997) and by Gillhooly and colleagues (Gillhooly et al., 1999; Phillips et al., 1999). However, specific results from this task cannot be directly compared with those from the Shallice TOL task, nor did Ward

[3]With further reduction of the open space to ball ratio the game becomes unplayable.

and Allport apparently intend them to be. We recommend that this task be simply described as the Ward and Allport tower task, dropping the Tower of London designation, in recognition of its being a unique task with its own characteristics that differ from both the TOH and the TOL.

In sum, we recommend that the Shallice TOL be regarded as the standard tower executive functioning task due to the wealth of literature available on this form of the TOL and its use with a wide range of ages and clinical populations. By carefully selecting problems, one need not be concerned with ceiling or floor effects with participants anywhere from 4 years of age upto old age, and with education levels from preschool up through college (Hughes, 1998; Krikorian et al., 1994; Rainville, Fabrigoule, Amieva, & Dartigues, 1998). However, if a researcher or clinician chooses to use a modification of the TOL, possibly to tailor the task for use on a specific cognitive process or clinical group, or to compare their results with one of the described variants of the TOL, we suggest that the researcher be explicit in the changes made to the apparatus used and clearly describe the problems administered. In this case, the researcher or clinician should be aware of the cost that problem modification may have: the lack of comparability to the larger body of TOL research. If a modification is warranted, then maintaining the 1:1 ratio of balls to open spaces will maximize comparability to the Shallice TOL versions since there would be a similar restriction on available paths to an optimal solution.

Computerized versions of the TOL task have benefits because of their ease of presentation and use, and their accuracy of data acquisition. However, computer presentation has been found to influence children's TOH performance (Bishop, Aamodt-Leeper, Creswell, McGurk, & Skuse, 2001). Therefore, to maximize compatibility with research involving a physical (e.g., wood or plastic) apparatus of the Shallice TOL task, we recommend that computerized versions of the task incorporate two features: (1) that the monitor image of the apparatus be as similar in appearance to the typical physical test apparatus as possible, and (2) that the movements necessary to reposition the balls be similar to those required when using the typical physical test apparatus. This second suggestion would allow for closer comparisons with the movement time to data derived from a physical test apparatus. We suggest that either a touchscreen be used (especially for very young or disabled participants) or that a mouse be employed to drag and drop the balls from one peg to the next. Virtual versions that move the ball by a simple sequence of two keystrokes, mouse clicks, or touches to a touchscreen are further removed from the actions required with the typical physical apparatus, making comparisons of movement times even more difficult. Evidence from TOH research (Svendsen, 1991) demonstrates that even using the same ball movement protocol but doing it with a mouse versus keystrokes can alter performance. Research directed at evaluating the impact of TOL problem presentation format is scarce and clearly needed.

ASSESSMENT OF PERFORMANCE ON THE TOL

One of the most significant difficulties in comparing TOL studies is due to the use of widely differing TOL performance measures. Part of this problem is endemic to any body of literature with a relatively new assessment tool. Investigators, both clinical and research oriented, need the flexibility to develop measures that most effectively evaluate the ability or cognitive construct of interest. However, a few investigators have primarily limited their assessment of TOL to a single measure of performance (e.g., Krikorian et al., 1994), and some published assessment procedures have used multiple performance measures but combined them into a single score (e.g., NEPSY).

Although single measures or rationally combined measures can be expedient in a clinical situation, they risk loss of critical information if they are not carefully considered. Thus, it is important to investigate multiple measures to determine which of them are most useful and informative as well as the optimal ways to combine them. There is evidence that some commonly used response measures are not highly correlated nor are they representative of the same underlying psychological construct (Byrd & Berg, in preparation; Levin, et al., 1996). It is important to note that, though the validity and reliability of these measures we are about to review have yet to be fully examined, the approach of using multiple performance measures allows the investigator to evaluate possible convergent results regarding a single problem-solving construct. Such an approach can begin to separate out possible independent components or subcomponents of problem solving.

We provide here preliminary recommendations for the use of individual response measures and suggestions for their further development. The performance measures reviewed below are categorized into four broadly defined groups that have been frequently represented in the literature: (1) success or accuracy of solution, (2) efficiency of the solution, (3) speed of performance and planning during the solution, and (4) rule breaks during the solution. In addition, we examine a set of measures used less often but with considerable potential for extending the information available from the TOL task and for differentiating between clinical and developmental groups. These categories are neither exhaustive nor completely independent of one another. However, we hope they may serve as a valuable heuristic and a starting point for examining the possibilities for their independent contribution to overall performance on the TOL task.

Success or Accuracy of Solution

We define ‘‘success measures’’ as measures that reflect the participants’ success in rearranging the balls from the starting position to the goal position. This would be reflected in the proportion or percent of the problems solved. Most investigators include some version of this measure, often referring to it as an ‘‘accuracy’’ measure.

Specific definitions of success measures are, in part, determined by the test procedure. One procedure limits participants to solutions reached in the minimum number of moves (sometimes called ‘‘perfect solutions’’) and the other procedure limits the time participants are allowed for their solution, but allows moves beyond the minimum needed. In the former case, definitions of success are based on the percentage solved on either the first attempt (e.g., Anderson et al., 1998; Luciana & Nelson, 1998) or in a fixed number of additional attempts (e.g., Levin, et al., 1996). A second approach defines success as the percent of problems solved within the time limit regardless of the number of extra moves needed to achieve that solution (e.g., Berg & Byrd, in preparation; Culbertson & Zillmer, 1998a, 1998b, 1998c). In some studies it is unclear what restriction is placed on the definition of success (e.g., Carnoldi, Barbieri, Gaiani, & Zocchi, 1999; Houghton et al., 1999).

One advantage of requiring solutions within the minimum number of moves is that it focuses participants on the need for careful planning. On the other hand, an important advantage of solutions restricted only by time rather than minimum moves is the ability to evaluate patterns of solutions, errors, and recovery from errors. Further, this latter approach may be maximally flexible since one can calculate success rates for occurrences of both perfect and less than perfect solutions. The approach of using a restricted time and the approach of using multiple tries at solution both allow for the evaluation of the efficiency of solution.

Efficiency of the Solutions

Problem-solving efficiency using the perfect solution approach is straightforward. The more attempts that are required to achieve a perfect solution, the less efficient is the solution; number of attempts is the measure of efficiency. Using the fixed time approach, problem-solving efficiency, as we define it, is the number of moves taken to reach a solution. Measures of this efficiency can be calculated in several different ways including counting the number of moves made on each problem, or, perhaps more preferably, counting the number of moves made beyond the minimum number of moves necessary for an optimal solution. Calculating the number of moves taken relative to the minimum moves allows one to compare efficiency on problems that differ in minimum moves.

More elegant versions of this efficiency measure may be possible if one makes a correction for the increasing chance of deviating from the optimal path as the number of minimum moves increases. This correction would involve an evaluation of the problem space to determine, at every step of the perfect solution, the number of available nonoptimal legal moves. One could calculate from this the overall probability of making a nonoptimal choice and use this as a correction in determining response efficiency. Corrected in this way, efficiency would be independent of the minimum moves necessary to perform a perfect solution and also independent of the increasing chance of deviant move errors.

Speed of Performance and Planning During the Solution

A common finding in tasks employed by cognitive psychologists is that speed of cognitive processing has a trade-off function relating it to accuracy of performance (e.g., Dennis, & Evans, 1996; Osman et al., 2000). Planning tasks also follow this same trade-off function, both in normal and frontally-lesioned adults (Davies, 2000; Luria, 1966). Thus, an obvious complement to any set of accuracy measures of the TOL is a measure of speed. Speed measures have been popular among TOL investigators (e.g., Culbertson & Zillmer, 1998a, 1998b; Luciana & Nelson, 1998; Morris et al., 1993). Unfortunately, there are limitations on the precision of speed measures when the task is manually administered and times are recorded via stopwatch (e.g., Anderson et al., 1996; Cockburn, 1995; Culbertson & Zillmer, 1998a, 1998b, 1998c and standardized batteries such as the NEPSY). However, when computerized versions of the tasks are employed (e.g., Byrd & Berg, in preparation; Owen et al., 1990) or timing analysis is done from videotape, off line, then significantly more accurate timing becomes achievable.

The first and most obvious of the speed measures is total solution time, obtained by recording the time from when the problem is first presented until the solution is finally achieved. This measure is reported to be sensitive to problem difficulty (Baker et al., 1996) and has also been found sensitive to severe head injury (Cockburn, 1995). One concern with the measure of the total solution time is that it combines multiple timing components that may well index quite different cognitive processes. The two primary timing components that we shall describe here we refer to as first move time, the time from the problem presentation to the completion of the first move, and subsequent move time, the time from the end of the first move until a solution is achieved. As described below, these two appear related to very different cognitive constructs, so we recommend that at a minimum the total time measures be broken down into these separate timing measures which, if so desired, can always be combined to obtain total move time (Luciana & Nelson, 1998).

First Move Time

The time from the problem presentation to the completion of the first move has been called "first move time," "planning time," "first move latency," or "initiation time." This period is comprised of the time taken to plan all or part of a solution and the time taken to actually reposition the ball once planning is completed; that is, a cognitive component and a motor component. With normal adults, the time to reposition the ball is usually far shorter than initial planning time. As a result, using first move time as a measure of planning time is not seriously contaminated by the motor component.

This initial motor component could be eliminated by ending the measurement when ball positioning was *initiated* rather than completed. However, participants will sometimes extend the ball motion time by briefly holding the selected ball above the pegs or moving the ball quite slowly toward its destination. These pauses and delays may occur because of continued planning or error checking during ball repositioning. Because the time involved in completing the ball motion can include continued cognitive processes, we recommend that the measurement of first move time continue until the *completion* of the actual ball repositioning.

It is important to note that if evaluating groups such as clinical patients, children or others with limited motor ability, the portion of first move time related to actual ball positioning could be considerably larger than with adult control groups. In these cases, independently determining the time needed to reposition a ball and subtracting this time from the first move time would be important. One valuable procedure to obtain ball reposition time is a yoked motor task (e.g., Owen et al., 1990).

If factors such as motor ability and task awareness are controlled, examination of first move time allows us to address either the extent to which the participant takes time to carefully consider a plan before initiating a solution or whether the participant can quickly develop an effective plan. Based on a factor analysis, Levin et al. (1996) report that first move time loads heavily on a different factor than do the measures of percent solved and number of broken rules, suggesting the importance of separately considering this measure. Unfortunately, a serious interpretative problem arises in considering just what aspect of planning is represented by the first move time.

One possible interpretation of first move time is that it becomes longer as participants carefully plan their future moves; in this case longer first move time would indicate more deliberative and thorough planning. This was the view of Levin et al. (1996) in explaining the factor analysis results of their study of head injured children.[4] Similarly, Kafer and Hunter (1997) reported longer planning (first move) times with more difficult problems, suggesting that first move time could be longer because a participant is forming a longer plan. However, Kafer and Hunter also suggest that first move time may be longer when a participant is struggling longer to formulate a plan. In these cases the longer first move time may not necessarily indicate any particular deliberateness or thoroughness, but possibly ineffective planning. We see evidence of this when participants clearly shorten their first move time as they gain experience, and presumably become more effective with the TOL task, even with problem difficulty held constant (Berg & Byrd, in preparation).

A correction for the number of moves in the problem as well as comparison of first move time with other measures, such as a success or efficiency, can help clarify whether group or individual differences in first move time are due to deliberateness or ineffective planning. Ward and Allport (1997) provide an example of the value of comparing first move time to other performance measures. These researchers found an inverse relationship between first move time and solution errors. These results make clear the need for separately scored, multiple measures of TOL performance to clarify the sources of differences in performance.

Subsequent Move Time

As noted above, total solution time can be divided into first move time and the subsequent time needed to solve the problem. The latter has been called "subsequent move time," "solution time," or "move time." This measure may serve as an index of problem-solving/planning speed when comparing groups with equal motor abilities. However, as with first move time, the subsequent time measure contains a combination of the motor time taken to reposition the balls during the solution, and the cognitive time taken for additional on-line planning and error correction. Because subsequent move time typically involves proportionally more motor time than does first move time, it is particularly valuable to separate motor and cognitive components with subsequent move time.

Researchers have used a number of ways to measure on-line planning pauses independent of ball repositioning. To compare groups of individuals that might have differed on motor speed, Owen et al. (1990) employed their computerized "One Touch Tower of London" and subtracted move times during a yoked motor control task from the move times during a problem solution. This resulted in a more pure measure of planning time during subsequent moves and allowed the authors between planning times in Parkinson's diagnosed and control adults.

Our laboratory has taken a different approach to the comparison of subsequent time planning in groups that differ on motor speed, an approach which does not require additional testing time needed for yoked motor controls. In comparing children and adults' TOL performance, we examined the time between each of a participant's

[4]The factor analysis of Levin et al. (1996) results in two aspects of planning. There is both a "planning and execution" factor, labeled "planning" in their Table 3, and a "planning and deliberation factor" which is labeled "inhibition" in that table. First move time loads heavily onto the "inhibition" factor.

moves for instances of unusually long pauses, reasoning that these delays usually occur because of on-line planning. Thus far, we have found that both the number of planning pauses and the proportion of planning pauses differentiates between the performance of children and adults (Byrd & Berg, under review). However, it is important to note that thus far we have used this procedure only as a measure of the frequency of planning pauses, not of the length of planning pauses.

Past research with the Tower of Hanoi by Spitz and colleagues has demonstrated that the simple measure of median move time may serve when comparing groups unlikely to differ in motor abilities (Spitz, Minsky, & Bessellieu, 1985). However, in studies comparing groups likely to differ in motor ability, such as groups of differing age or clinical versus nonclinical groups, the within-subject based scores or the yoked motor control procedure described above may be useful in correcting for differences in motor speed or motor initiation.

Rule Breaks During the Solution

The Shallice TOL includes three rules: (1) do not place more than the allowed number of balls on any peg, (2) do not place the balls anywhere other than on a peg, and (3) only reposition one ball at a time. During solutions, participants may attempt to break any one of these rules. In past literature, the instances of this behavior have been handled in a wide variety of ways. Some studies, including those using the currently available computerized tests, do not measure instances of this behavior and do not allow the illegal move to be made. A few studies with clinically diagnosed patients, such as children with closed head injury and adults with dementia, have measured a combined score for all types of rule breaks, and have found it to be more frequent in clinical groups (Carlin et al., 2000; Levin et al., 1997). To our knowledge, different types of rule breaking either have not been scored separately, or have been scored separately but then combined for the reported analysis (e.g., Culbertson & Zillmer, 1998b).

A concern that comes up in the interpretation of rule breaking scores comes from the fact that a rule break could occur for two reasons. Either the participants are unable to remember the rules or, alternatively, participants know the rules but choose to disregard them. In the literature on neurologically disabled adults, it seems implicit in the articles that rule breaks occur because participants forget the rules or have problems acting within those rules in the task. However, when one is testing normal participants, children or adults, it may be more difficult to interpret the cause of rule breaking behavior. To help in interpreting measures of rule breaking with any population, we recommend procedures that insure participants understand the rules, such as the use of practice problems or the use of a series of questions that requires the participant to explain the rules or identify illegal moves to test personnel.

Measures Tapping New Sources of Information

Gaze Focus During the Solution

When solving a TOL problem, the participant has two components of the task that she or he must visually attend to: the game board, where the balls are manipulated to solve the task, and the goal board, which shows the arrangement of balls that the participant is attempting to match. It has been found that participants attend to these two parts of the task at different times during their solution. Hodgson, Bajwa, Owen, and Kennard (2000) found that adult participants initially look at the goal position, but then as they proceed through problem solving, better solvers look almost exclusively at the game board. This measure of duration and placement of gaze, to our knowledge, has only been used in this one study with normal adults. However, the measure of gaze direction offers promise as a measure of ability to maintain the goal position in memory and to sustain attention to the task, either of which may differ between clinical or developmental groups.

Task Exploration and Solution Execution

Kotovsky et al. (1985) found that normal adults' solutions on the TOH task contain two phases indicative of two different cognitive processes. They refer to the first phase as ‘‘exploratory behavior’’ where the participants make seemingly random moves, some directed closer to the goal and others further away. Research by Zhang (1997) suggests that during this phase, participants are forming a mental representation of the problem. The second phase noted by Kotovsky, Hayes, and Simon is referred to as "final path behavior." This phase occurs when participants recognize the final sequence of moves that they must take to achieve a solution and then "close in on the goal," directly and unerringly.

The conceptualization of two phases of TOH solutions can be adapted from the work of Kotovsky et al. (1985) and applied to the TOL, provided that the TOL test procedure allows solutions other than perfect (optimal) solutions. Each move of TOL solution can be analyzed within the problem space to determine if it has taken the participant closer or further from the goal. This "distance to goal" measure can be calculated at the completion of each move and based on the minimum number of moves to reach the goal from the current position.

The length of the first phase of problem solution, the Pre-Goal Phase, is defined as the number of moves made prior to initiating the Goal Phase. It begins with the first move of the solution and ends with the first move of the Goal Phase. In contrast to the Goal Phase, moves during the Pre-Goal Phase put the ball closer to or further from the goal. The minimum Pre-Goal Phase length is zero, which can occur when there is an optimal solution since, by definition, this solution takes the participant unerringly to the goal position. In this circumstance the entire solution is comprised of the Goal Phase. The maximum Pre-Goal Phase length is limited only by the number of moves possible in the time allotted for solving a problem. A shorter Pre-Goal Phase length may indicate that fewer moves are required to develop an accurate solution plan and therefore an earlier focus on reaching the goal.

The Goal Phase is the sequence of moves that lead to the solution wherein each move brings the participant one move closer to the goal. The Goal Phase length has the possible range of 1 to 8 moves. The maximum is 8 since in the Shallice TOL problem space one can never be more than 8 optimal moves from any other position. For optimally (perfectly) solved problems, the Goal Phase is always equal to the minimum numbers of moves for the problem being solved. However, with nonoptimal solutions the Goal Phase length may exceed the minimum number of moves of the problem since the participant may have moved further from the goal during the Pre-Goal Phase. In general, a longer Goal Phase may be indicative of a plan that involves a deeper solution search; that is, a plan that considers the consequences of a greater number of moves in advance of the current position.

Kotovsky et al. (1985) and Zhang (1997) argue that the two phases of problem solving reflect very different cognitive processes. As such we might expect that developmental and clinical differences in planning and problem-solving abilities may be reflected in the relative lengths of these two separate phases of problem solution in the TOL. To our knowledge, our laboratory alone has examined this question with the TOL. Thus far, we have found the phases to differ between children and adults' solutions (Byrd & Berg, under review) as well as being sensitive to testing manipulations in a group of young children (Byrd, van der Veen, Berg, & McNamara, under review).

Learning Rate

One category of performance that, surprisingly, has been largely unused with the TOL task is the amount or pattern of improvement in task performance over a series of problems. Success scores averaged over a series of problems incorporate both initial performance and any change in performance over the series. It is possible that these two components could be independent of one another, which can only be determined from separate analysis of each. Learning rate performance on the TOL task may provide a useful clinical and cognitive index of ability to increase planning effectiveness with experience and is clearly in need of research.

Learning rates can be assessed in a variety of ways. When participants are required to make minimum-move solutions, several investigators have measured learning rates as reductions in the number of attempts needed to obtain this optimal solution (e.g., Anderson, et al., 1996; Krikorian, et al., 1994; Owen et al., 1995). In the typical procedure associated with obtaining this measure, as soon as the solver makes the minimum number of moves without reaching a correct solution, the task is interrupted, and the same problem is begun again. This re-presentation procedure is a substantial change from the single problem presentations used originally by Shallice, and could potentially cause frustration in children, in clinical populations, or with the presentation of difficult problems. Another potential downside of this procedure is that it prevents the measure of solving efficiency discussed above. However, this re-presentation procedure may be valuable in its emphasis on the careful planning of an entire, efficient solution. One possible modification of this re-presentation procedure would be to allow the participant to complete a nonoptimal solution with extra moves, before presenting the task again. This would, of course, lengthen the entire procedure, but the benefits may well outweigh the costs in instances where both learning and efficiency measures are of interest.

A more obvious measure of learning rate that, to our knowledge, has not been explored is the straightforward improvement of performance across problems. Faster rates of learning (steeper learning curves) are presumed to be related to more effective learning from experience with the TOL. Perhaps the reason this measure of performance has not been utilized is that participants typically receive only 12 to 15 problems, and problems are usually presented in increasing order of difficulty, making it difficult to assess learning rates. Our experience has been, however, that even with very few trials and in the face of increasing problem difficulty there is evidence of learning (Berg & Byrd, in preparation). Additional information can be obtained when a longer series of problems can be presented, providing information on changes in planning as problem-solving expertise evolves.

Incomplete Moves During the Solution

During solutions, participants may pick up a ball from the peg and then return it to the original position, without having placed it on another peg. The measure of the number of these moves may reflect impulsivity or lack of planning. Specifically, we have found this behavior to be more common in children's solutions as compared to adults' solutions (Byrd & Berg, under review). It appears that participants pick up the ball without having planned out a full sequence of moves and then return the ball as they do form a plan or change their plan. To our knowledge, no research has examined this measure with clinical patients. Since our laboratory's work with this measure is still preliminary and there is a paucity of published work with this measure of incomplete moves, this measure should be used on an exploratory basis.

RECOMMENDATIONS, FUTURE DIRECTIONS, AND CONCLUSIONS

The rapid expansion of the application of the TOL since its development by Shallice (1982) certainly appears to be warranted. This task has provided a wealth of information and appears to contain promise for further development and refinement. However, the true promise of this task in the exploration of executive functions and evaluation of neuropsychological conditions will not be realized without a more complete understanding and utilization of this research and clinical tool. Without the careful use of this, or any, tool we risk confusion, loss of vital information, and a reduced ability to compare results across studies. Toward this end we provide a summary of our major recommendations, and some promising directions for future work.

Recommendations

1. Report specific starting and goal positions for every problem in a study. For the Shallice TOL we suggest the numbering system space described above based on the problem space (Fig. 2).
2. When utilizing a variant of the TOL, consider using the naming system suggested above and a clear description of the apparatus, testing procedures, and problem selection so as to allow readers to ascertain the similarity of the variant to the Shallice TOL.
3. Use multiple measures of TOL performance. At a minimum, there should be at

least one speed measure and one accuracy measure.

a. We suggest the use of first move time and planning pauses as valuable speed measures. Also, it is far more informative to report first move time and subsequent move time separately as compared to only total solution time.
b. We recommend the use of percent correct solutions as the success measure along with a testing procedure that allows participants to continue attempting to solve a problem, even when the minimum moves for a problem have been exceeded. This allows for detailed examination of solution efficiency and errors.
c. Combined performance measures are clinically useful, but should not be used without providing the component scores as well, in order to facilitate cross-study comparisons. Optimally, a combined score should be based on a psychometrically sound integration and weighting of basic performance measures.

Future Directions

1. There is a clear need for further factor analytic and other multivariate assessments of a variety of performance measures on a large normative population. The focus should be on selecting the independent contributors of TOL performance, weighting their value, and rationally developing a combined measure or measures.
2. Clearly researchers would benefit from expanding the literature concerning learning rates for the TOL task, and for the differences in task approaches between more experienced and novice participants.
3. There is a need for basic, empirical information on the effects of changes in the method of task presentation (e.g., computer interface devices). For example, what are our losses and gains as we move from testing with a physical test apparatus to testing with a computer simulation, and what are the differences between using the touchscreen versus the mouse to interface with a computer version of the TOL? If we are to compare across studies and effectively use normative data, these seemingly mundane issues must be addressed.
4. There is a need for further research into the validity and reliability of individual measures described above in both clinical and developmental populations. Both Culbertson and Zillmer (1998c) and Schnirman et al. (1998) provide some of this information.
5. There is a need for standardized procedures and normative data on the individual measures of performance described above. The standardized data available are largely for composite measures of performance (e.g., NEPSY; Anderson et al., 1996; Krikorian et al., 1994), although Culbertson and Zillmer (1998b, 1998c) report norms for some individual measures.

CONCLUSIONS

We have considered in this paper the basic aspects of the TOL task including its problem space, current research, variants, performance measures, and recommendations for future use. It is clear from this examination that the TOL has gained a very strong foothold in the research and clinical communities, deservedly so. The strength of this task, however, is currently limited by the wide variations in use, misunderstandings regarding the impact of these variations, and a need for the rational choice of performance measures. Part of the reason for these limitations is the lack of an adequate groundwork with nonclinical populations. The TOL has not been explored fully as a cognitive tool by researchers in the way that the TOH has. Thus, we conclude with a call for more basic research on the TOL by both clinical and cognitive researchers with a clear eye toward the TOL's powerful clinical applications.

ACKNOWLEDGMENTS

We are indebted to Dr. Steven Blessing for seminal ideas regarding the problem space and to the undergraduate and graduate assistants in our laboratory who have contributed significantly to the refining of the ideas expressed in this article.

REFERENCES

Anderson, P., Anderson, V., & Lajoie, G. (1996). The Tower of London Test: Validation and standardization for pediatric populations. *The Clinical Neuropsychologist*, *10*, 54–65.

Baker, S.C., Rodgers, R.D., Owen, A.M., Frith, C.D., Dolan, R.J., Frackowiak, R.S.J., & Robbins, T.W. (1996). Neural systems engaged by planning: A PET study of the Tower of London task. *Neuropsychologia*, *34*, 515–526.

Berg, W.K., & Byrd, D.L. (2001). Parameters of problem difficulty in the Tower of London task. In preparation.

Bishop, D.V., Aamodt-Leeper, G., Creswell, C., McGurk, R., & Skuse, D.H. (2001). Individual differences in cognitive planning on the Tower of Hanoi task: Neuropsychological maturity or measurement error? *Journal of Child Psychology and Psychiatry*, *42*, 551–556.

Byrd, D.L., & Berg, W.K. (under review). Getting to the heart (rate) of planning: Developmental differences in the use of a planning period. *Developmental Psychology.*

Byrd, D.L., van der Veen, T., Berg, W.K., & McNamara, J.P. (under review). Requirements of speech and motor movement differently impact preschoolers' problem solving performances. *Journal of Cognition and Development*.

Carlin, D., Bonberba, J., Phipps, M., Alexander, G., Shapiro, M., & Gagman, J. (2000). Planning impairments in frontal lobe dementia and frontal lobe lesion patients. *Neuropsychologica*, *38*, 655–665.

Carnoldi, C., Barbieri, A., Gaiani, C., & Zocchi, S. (1999). Strategic memory deficits in attention deficit disorder with hyperactivity participants: The role of executive processes. *Developmental Neuropsychology*, *15*, 53–71.

Cockburn, J. (1995). Performance on the Tower of London test after severe head injury. *Journal of the International Neuropsychological Society*, *1*, 537–544.

Culbertson, W.C., & Zillmer, E.A. (1998a). The construct validity of the Tower of London^{DX} as a measure of the executive functioning of ADHD children. *Assessment*, *5*, 215–226.

Culbertson, W.C., & Zillmer, E.A. (1998b). The Tower of London^{DX}: A standardized approach to assessing the executive functioning in children. *Archives of Clinical Neuropsychology*, *13*, 285–301.

Culbertson, W.C., & Zillmer, E.A. (1998c). *The Tower of London^{DX} Manual*. Multi-Health Systems, Inc. (Available at www.mhs.com).

Dagher, A., Owen, A.M., Boecker, H., & Brooks, D.J. (1999). Mapping the network for planning: A correlational PET activation study with the Tower of London task. *Brain*, *122*, 1973–1987.

Davies, S.P. (2000). Memory and planning processes in solutions to well-structured problems. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *53A*, 896–927.

Dennis, I., & Evans, J. St. B.T. (1996). The speed-error trade-off problem in psychometric testing. *British Journal of Psychology*, *87*, 105–129.

Gillhooly, K.J., Phillips, L.H., Wynn, V., Logie, R.H., & Della Sala, S. (1999). Planning processes and age in the five-disc Tower of London task. *Thinking and Reasoning*, *5*, 339–361.

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.

Houghton, S., Douglas, G., West, J., Whiting, K., Wall, M., Langsford, S., Powell, L., & Carroll, A. (1999). Differential patterns of executive function in children with Attention-Deficit Hyperactivity Disorder according to gender and subtype. *Journal of Child Neurology*, *14*, 801–805.

Hughes, C. (1998). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology*, *34*, 1326–1339.

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.

Klahr, D. (1985). Solving problems with ambiguous subgoal ordering: Preschoolers' performance. *Child Development*, *56*, 940–952.

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

Kotovsky, K., Hayes, J.R., & Simon, H.A. (1985). Why are some problems hard? Evidence from the Tower of Hanoi. *Cognitive Psychology*, *17*, 248–294.

Lange, K.W., Robbins, T.W., Marsden, C.D., James, M., Owen, A.M., & Paul, G.M. (1992). L-Dopa withdrawal in Parkinson's disease selectively impairs cognitive performance in tests sensitive to frontal lobe dysfunction. *Pharmacology*, *107*, 394–404.

Levin, H.S., Fletcher, J.M., Kufera, J.A., Harward, H.J., Lilly, M.A., Mendelsohn, D., Bruce, D., & Eisenberg, H. (1996). Dimensions of cognition measured by the Tower of London and other cognitive tasks in head-injured children and adolescents. *Developmental Neuropsychology*, *12*, 17–34.

Levin, H.S., Song, J., Scheibel, R.S., Fletcher, J.M., Harward, H., Lilly, M., & Goldstein, F. (1997). Concept formation and problem-solving following closed head injury in children. *Journal of the International Neuropsychological Society*, *3*, 598–607.

Luciana, M., & Nelson, C.A. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologica*, *36*, 273–293.

Luria, A.R. (1966). *Higher cortical functions in man*. New York: Basic Books, Inc.

MacDonald, C., Garner, E., & Spurgeon, J. (in press). Executive functioning and planning abilities of preschool children using the Tower of London task. *Psi Chi Journal of Undergraduate Research*.

Morris, R.G., Ahmed, S., Syed, G.M., & Toone, B.K. (1993). Neural correlates of planning ability: Frontal lobe activation during the Tower of London test. *Neuropsychologia*, *31*, 1367–1398.

Morris, R.G., Downes, J.J., Sahakian, B., Evenden, J.L., Heald, A., & Robbins, T.W. (1988). Planning and spatial working memory in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, *51*, 757–766.

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.

Osman, A., Lou, L., Muller-Gethmann, H., Rinkenauer, G., Mattes, S., & Ulrich, R. (2000). Mechanisms of speed-accuracy tradeoff: Evidence from covert motor processes. *Biological Psychology*, *51*, 173–199.

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

Owen, A.M., Sahakian, B., Hodges, J.R., Summers, B.A., Polkey, C.E., & Robbins, T.W. (1995). Dopamine-dependent frontrostriatal planning deficits in early Parkinson's disease. *Neuropsychology*, *9*, 126–140.

Phillips, L.H., Wynn, V., Gilhooly, K.J., Della Sala, S., & Logie, R.H. (1999). The role of memory in the Tower of London task. *Memory*, *7*, 209–231.

Rainville, C., Fabrigoule, C., Amieva, H., & Dartigues, J.F. (1998). Problem-solving deficits in patients with dementia of the Alzheimer's type on a Tower of London task. *Brain and Cognition*, *37*, 135.

Rezai, K., Andreasen, N.C., Alliger, R., Cohen, G., Swayze, V., & O'Leary, D. (1992) The neuropsychology of the prefrontal cortex. *Acta Neurologica*, *50*, 636–642.

Schnirman, G.M., Welsh, M.C., & Retzlaff, P.D. (1998). Development of the Tower of London – revised. *Assessment*, *5*, 355–360.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London*, *Biology*, *298*, 199–209.

Shallice, T., & Burgess, P. (1991). Higher-order cognitive impairments and frontal lobe lesions in man. In H.S. Levin, H.M. Eisenberg, & A.L. Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 125–138). New York: Oxford University Press.

Spitz, H.H., Minsky, S.K., & Bessellieu, C.L. (1984). Subgoal length versus full solution length in predicting Tower of Hanoi problem-solving performance. *Bulletin of the Psychonomic Society*, *22*, 301–304.

Spitz, H.H., Minsky, S.K., & Bessellieu, C.L. (1985). Influence of planning time and first-move strategy on Tower of Hanoi problem-solving performance of mentally retarded young adults and nonretarded children. *American Journal of Mental Deficiency*, *90*, 46–56.

Svendsen, G.B. (1991). The influence of interface style on problem solving. *International Journal of Man-Machine Studies*, *35*, 379–397.

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

Welsh, M.D., 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.

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*, *21*, 179–217.