The development and preliminary validation of the Tower of Hanoi-Revised

Welsh, M.C.; Huizinga, M.

DOI
10.1177/107319110100800205

Publication date
2001

Document Version
Final published version

Published in
Assessment

ASSESSMENT
2001, Volume 8, Number 2, 167-176

THE DEVELOPMENT AND PRELIMINARY VALIDATION OF THE TOWER OF HANOI–REVISED

Marilyn C. Welsh
University of Northern Colorado

Mariëtte Huizinga
University of Amsterdam

The objective of this study was to redesign the Tower of Hanoi (TOH) in order to increase the reliability and explore the validity of the task. In order to achieve this goal, a two-phase process was undertaken. In Phase 1, the TOH item pool was enlarged from 12 items to 60 items, and the task was administered to a sample of college students (*N* = 81). The 22 most homogeneous items for the final task were identified based on the item-total correlations. In Phase 2, the rebuilt TOH (TOH-R) was administered to a second sample (*N* = 50). The internal consistency reliability (Cronbach alpha) was .77 for this new sample. In this phase, the TOH-R was administered concurrently with the Tower of London–Revised (TOL-R), and the convergent validity of the TOH-R against the TOL-R was .53.

Keywords: Neuropsychology, psychometrics, Tower of Hanoi, executive function, Tower of London

The cognitive capacity to regulate behavior and more basic cognitive processes to achieve a future goal is referred to as executive function (Welsh, 1991; Welsh & Pennington, 1988). Executive function has been hypothesized to use mechanisms of goal selection, planning, set maintenance, self-monitoring, inhibition, and flexibility of strategies (Shallice, 1982, Stuss & Benson, 1984; Welsh & Pennington, 1988). The prefrontal cortex presumably mediates executive processes (Stuss, 1992), which initiate and monitor comparisons between goals and obtained results by using a temporal series of reference points (Welsh & Pennington, 1988). Therefore, executive functioning allows self-monitoring of complex, goal-directed activity, particularly in novel situations (Brennan, Welsh, & Fisher, 1997; Stuss, 1992; Welsh, Pennington, & Groisser, 1991). The Tower of Hanoi (TOH; e.g., Simon, 1975) and the Tower of London (TOL; Shallice, 1982) are considered to be exemplar measures of executive function. These tests require the ability to cope adequately with novel situations as well as anticipatory, means-end problem-solving demands.

The TOL and the TOH are examples from a class of problem-solving tasks often referred to as disk-transfer tasks. These tasks are similar in the sense

Marilyn C. Welsh, Department of Psychology, University of Northern Colorado, Mariëtte Huizinga, Department of Developmental Psychology, University of Amsterdam, Netherlands.

The authors wish to acknowledge Molly McGarraugh, Sarah Devine, Kasey Beaty, Christy Adams, and Casey Ryan for their invaluable assistance during the data collection phases of this project. We also gratefully acknowledge the comments of Dr. Paul Retzlaff on an earlier draft of this paper.

Correspondence concerning this article and requests for offprints should be addressed to Marilyn Welsh, Department of Psychology, University of Northern Colorado, Greeley, CO, 80639. E-mail: mcwelsh@bentley.unco.edu

that they require the transformation of a start state of balls or disks on three vertical pegs into a goal configuration of these objects in the fewest number of moves. Both the TOH and the TOL impose a set of rules that constrain the manner in which these objects may be moved from peg to peg. Given that the most efficient solution is demanded for both tasks, a sequence of moves must be planned, executed, monitored, and revised prior to action. Although the TOL and the TOH share the general properties and demands of disk-transfer tasks, these tasks also vary substantively in structure, administration, and evaluation of performance (Welsh & Pennington, 1992).

In clinical and experimental neuropsychology, the TOH and the TOL are considered to be essentially isomorphic. That is, it is assumed that the two tasks are versions of the same general task and, thus, assess the same set of cognitive processes. One line of evidence consistent with this view is the sensitivity of each task to frontal cortical damage. Recent research suggests that performance on TOL and TOH is sensitive to prefrontal damage or dysfunction (e.g., Goel & Grafman, 1995; Glosser & Goodglass, 1990; Levin et al., 1994; Shallice, 1982). There is also evidence that performance on the TOL and the TOH is impaired in clinical groups with diagnoses thought to reflect prefrontal dysfunction as well as general cortical pathologies. For example, deficient performance on the TOH has been found in children diagnosed with attention deficit hyperactivity disorder (ADHD; Pennington, Groisser, & Welsh, 1993), children with early-treated phenylketonuria (PKU; Welsh, Pennington, Ozonoff, McCabe, & Rouse, 1990), women with fragile-X syndrome (Mazzocco, Hagerman, Cronister, & Pennington, 1992), high-functioning autistics (Ozonoff, Pennington, & Rogers, 1991), and schizophrenics (Goldberg, Saint-Cyr, & Weinberger, 1990). Similarly, impaired performance on the TOL has been observed in clinical groups such as individuals with Parkinson's disease (Hanes, Andrewes, Smith, & Pantelis, 1996; Lange et al., 1992; Owen, Downes, Sahakian, Polkey, & Robbins, 1990) and ADHD (Cornoldi, Barbieri, Gaiani, & Zocchi, 1999). Researchers have suggested that the TOL and the TOH are sensitive to frontal lobe functions because they tap the executive functions of working memory (Goldman-Rakic, 1987), inhibition (Goel & Grafman, 1995) or both cognitive processes (Roberts & Pennington, 1996). However, empirical support for the construct validity of these tasks is still lacking.

Recent research conducted by Humes, Welsh, Retzlaff, and Cookson (1997), and Welsh, Satterlee-Cartmell, and Stine (1999) has shown a substantial lack of overlap between the performance on the TOH and the TOL. Significant, but moderate, correlations between the performance on the TOH and the TOL were found (r ranges from .35 to .60). Moreover, these correlations suggest that approximately 75% of the variance in performance on the two tasks is not shared. This nonshared variance represents both unique true variance as well as the error variance associated with each task. This lack of shared variance may be the result of any one, or more, of the following three factors. First, different cognitive demands that may be posed by the TOH and TOL would contribute to the true variance unique to each tower task. Second, different administration and scoring procedures across the two tasks may contribute to both unique true variance and error variance, and these differences are listed in Table 1. Third, the potentially low reliability of one or both tasks would contribute to measurement error alone.

Regarding the first explanation, Welsh et al. (1999) explored the common assumption that both TOL and TOH measure working memory and inhibition processes. With regard to the TOL, inhibition and working memory scores significantly predicted overall performance. The combination of all the working memory and inhibition variables explained over half of the variance in TOL performance. However, there was a lack of significant predictors of TOH performance, and there was no significant correlation between performance on TOH and working memory variables. Only errors on one inhibition task and pure naming speed on the Stroop Color-Word Interference Task (Stroop, 1935) significantly predicted performance. This study provides preliminary evidence that the two tower tasks may be assessing different cognitive processes, as has been suggested elsewhere (Goel & Grafman, 1995).

The second possible explanation for the substantial amount of nonshared variance between TOH and TOL was tested by Welsh, Revilla, Strongin, and Kepler (2000). These researchers examined the degree to which some of the administration differences between the two tasks contribute to nonshared variance. A new one-trial version of the TOH (TOH-1) was developed to match the TOL in three procedural aspects: (a) participants were given only one trial to solve each problem; (b) participants were told, in advance, the number of moves required to solve each problem; and (c) all problems were given, regardless of performance. The main structural difference between TOH and TOL (i.e., disks vs. balls, pegs of different heights vs. disks of different sizes) are integral, characteristic features of each task that were not altered. The main question addressed in this study was the extent to which these three procedural changes would change the association between the two tasks. That is, if these procedural features were contributing to some degree to the nonshared variance between the TOH and the TOL, then changing the TOH to be identical to the TOL in these features should increase the inter-task correlation. The results of this study indicated intertask correlations in the same range as had been found previously (r = .40 – .60), and, therefore, administration differences were not responsible for the nonshared variance.

Importantly, one cannot fully address the contribution of factors, such as cognitive mechanisms and task administration differences, to the lack of shared variance between TOH and TOL without an understanding of the psychometric properties of each task. The third explanation for the nonshared variance rests on the fact that, if one or both tasks have low reliability, the degree to which they will correlate will be limited. There has been very little research to date on the psychometric integrity of the TOL (Culbertson & Zillmer, 1998; Humes et al., 1997; Schnirman, Welsh, & Retzlaff, 1998), and research designed to examine the psychometric properties of the TOH is even more rare. Recently, Humes et al. (1997) discovered that the TOL typically used in research (e.g., Levin et al., 1994) had a very low internal consistency (alpha = .25). In order to increase the internal consistency, Schnirman and colleagues (1998) modified the original TOL developed by Shallice (1982). The internal consistency of the newly designed 30-item TOL–Revised (TOL-R) increased to a Cronbach alpha of .79. Furthermore, a test-retest correlation of .70 was found. This internal consistency has since been replicated in our laboratory (Welsh et al., 2000).

As part of the Humes et al. (1997) study, the internal consistency of the TOH also was examined and found to be .91. However, based on the results of a subsequent study by Welsh et al. (1998), this high reliability appears to be the consequence of the unique administration features characteristic of the TOH. That is, the procedure in which testing is terminated at the first failed problem creates a local dependency among the item scores, inflating the internal consistency index. Additionally, the procedure of providing six trials per problem to determine an item score also contributes to this inflated index. When the TOH was redesigned to eliminate these procedures (i.e., TOH-1: all problems and *one* trial per problem were given), the alpha coefficient dropped to .40. Because the TOH consists of only 12 problems, the reason for the low internal consistency of the TOH-1 may be that this relatively small set of problems does not allow for the presentation of a variety of problem configurations of varying difficulty levels. That is, the findings of this study indicated a need for a greater number and variety of TOH problems to achieve a reliable assessment of underlying executive function processes.

The objective of the present study was to redesign the TOH, using the same procedures used in redesigning the TOL-R (Schnirman et al., 1998), in order to increase the reliability and explore the validity of the task. Classical test construction theory was employed in which the pool of available items was enlarged by varying the start and goal configurations. Next, this large set of problems was administered to a sample of normal college students, and the best predictors for the final task were identified. Theoretically, this step should yield a homogeneous item pool, consequently increasing the reliability of the task. Finally, the rebuilt TOH (TOH-R) was administered to a second

Table 1
Item-Total Correlations for the Individual Items of the 22-Item TOH-R, in Phase 1 and Phase 2

			Phase 1	Phase 2
Serial position	Number of moves	Goal state	Item-total correlation	Item-total correlation
1	7	Tower	.328	.308
2	8	Flat	.366	.049
3	8	Tower	.382	.342
4	9	Flat	.290	.460
5	11	Tower	.433	.383
6	11	Flat	.357	.330
7	12	Tower	.373	.502
8	11	Flat	.505	.338
9	12	Tower	.425	.166
10	12	Flat	.377	.410
11	12	Tower	.462	.338
12	13	Flat	.420	.148
13	13	Tower	.445	.461
14	14	Flat	.329	.441
15	13	Tower	.354	.082
16	14	Flat	.221	.112
17	13	Tower	.471	.299
18	14	Flat	.383	.345
19	15	Tower	.540	.350
20	15	Flat	.460	.260
21	15	Tower	.336	.310
22	15	Flat	.469	.430

Note. Item-total correlations for Phase 1 reflect the correlation of each item with the total score on the 60-item TOH-task. Item-total correlations for Phase 2 reflect the correlation of each item with the total score on the 22-item TOH-R task. TOH-R = Tower of Hanoi–Revised.

sample of college students in order to obtain an independent measure of internal consistency, an estimate of the test's reliability. In order to estimate the convergent validity of TOH-R, the task was administered concurrently with the TOL-R. The convergent validity of the newly designed TOH-R was explored by examining its correlation with the TOL-R. Given that the validation procedure employed two psychometrically sound tasks that were comparable in important administration features (see Table 1), much of the nonshared variance between the TOH-R and the TOL-R may be indicative of different cognitive processes underlying performance on each task.

Phase 1

Phase 1 involved the development of the revised TOH (TOH-R). First, the pool of items was increased from the original 12 to 60 and administered to a sample of college students. The 22 items that correlated most highly with the overall scores were retained. The goal was to increase the internal consistency of the TOH beyond that found by Welsh et al. (2000) on the TOH-1.

Method

Participants

A total of 87 undergraduate students enrolled at a midsized university participated in the study. Given that this study represented a first step to assess the basic psychometric properties of a newly designed test, our objective was to select a homogeneous sample of normally functioning college students. Therefore, if a participant reported a history of head injury and/or had a diagnosed learning disorder, that participant was excluded from the study. Six participants were excluded from the analysis for the following reasons: three participants (two females, one male) were excluded due to self-reports of a past head injury, and three participants (three females) were dropped from the analyses due to extremely low scores (well below 2 standard deviations from the sample mean) and behaviors indicating a failure to understand the task. Therefore, the sample of participants included 81 students, 60 females and 21 males, with a mean age of 18.32 years ($SD = .70$ years). The ethnic composition of the sample was not recorded; however, the sample was primarily White, mirroring the demographic makeup of the university. The students were compensated with course credit for their participation.

Apparatus

The Tower of Hanoi (TOH; Simon, 1975) consists of a flat board (40 x 15 x 2 cm) on which three vertical wooden pegs of equal diameter (1 cm) and equal height (14.5 cm) are spaced equidistantly (12.5 cm). Three or four wooden disks of graduated size (13.5, 11, 8.5, and 6 cm diameter) are also included, each disk has one hole (1.3 cm in diameter) drilled through the center so that it will fit onto any of the three pegs. A set of 17 x 22 cm cards in a three-ring binder displays the goal states of the individual items that are presented to the participant.

The TOH requires that an initial start configuration of disks across the three vertical pegs be transformed into a specific goal configuration of these objects in the minimum number of moves. Disks must be moved according to a set of specified rules that constrain the manner in which these objects may be moved from peg to peg. These rules include the following: (a) only one disk may be moved at a time; (b) a disk may not be placed on the table or held in the hand while another disk is being moved; and (c) a larger disk may not be placed on top of a smaller disk.

Procedure

In the current phase, the size of the item pool was increased. By varying the start and goal configurations, 60 four-disk TOH problems were generated. There were six items at each of 10 levels, representing 6-move through 15-move items. The task consisted of 30 tower-ending problems and 30 flat-ending problems. Participants were administered the TOH individually after an explanation of the three rules. For each item, the tester set up the start state on the TOH apparatus, and the participant was presented with a card that exhibited the goal state. Both this card and the tester indicated the number of moves required to reach the goal state. Participants had to reach the goal in the designated number of moves on the first attempt, and there was no time limit imposed. Scoring involved awarding one point for each correct solution (i.e., transforming the start state to the goal state in the required number of moves).

Results

The variable of interest was the number of items the participants solved correctly. This score is dichotomous (i.e., 0 or 1), indicating whether or not the participants solved the problem correctly.

In order to determine which of the items to retain for the TOH-R, the relationship of each individual item to the whole test was analyzed by means of item-total (point-biserial) correlations. These correlations ranged from .03 (a 10-move item) to .54 (a 15-move item). The 22 items with the highest item-total correlation were selected to compose the TOH-R for use in Phase 2. Of these 22 items, two 15-move problems (items #19 and #22) were part of the original TOH (eg., Humes et al., 1997). The range of item-total correlations for the reduced set was between .22 (a 14-move item) and .54 (a 15-move item). Table 1 presents these items in terms of the number of moves required for correct solution as well as the initial item-total correlation for each.

Based on these 22 items, the range of the total scores was 3 to 22 out of a possible 22, and the mean total score was 13.64 ($SD = 4.83$). These descriptive statistics do not suggest floor or ceiling effects and point to good resolution and range.

Phase 2

The purpose of this phase of testing was to obtain an independent estimate of reliability of the TOH-R. A concurrent purpose was to assess the convergent validity of the TOH-R with the TOL-R (Schnirman et al., 1998). Both tasks were administered to a second sample of normal college students.

Method

Participants

A total of 52 undergraduate students enrolled at a midsized university participated in this study. Students who participated in Phase 1 were not allowed to participate in Phase 2. Upon volunteering for the study, the participants were questioned as to known diagnosis of any learning disabilities or head injuries. If the participant had a history of head injury and/or had a diagnosed learning disorder, that participant was excluded from the study. Two participants were excluded from the analyses due to TOH-R scores below 2 standard deviations from the mean (i.e., scores of 1 and 2 correct of 22 total items). Therefore, the sample of participants included 50 students, 36 females and 14 males, with a mean age of 19.20 years ($SD = 2.02$ years). These students were compensated with course credit for their participation.

Apparatus

Two tasks were used in Phase 2, the TOH-R and the TOL-R (Schnirman et al. 1998). The TOH-R apparatus was described in Phase 1. The item pool consisted of 22 items drawn from the original 60 items (see Figure 1). The 30-item TOL-R used in this study was adapted from the task developed by Shallice (1982) and modified by Schnirman et al. (1998) to be more internally consistent (alpha = .79).

The TOL-R apparatus consists of a wooden base (18.5 x 6.5 x 2 cm) with three vertical wooden pegs of equal diameter (1 cm) and of differing heights (16.5, 11.5, 6.5 cm) attached to the middle of the base and equidistant from each other (5 cm). Three equal sized wooden balls (5 cm in diameter) of different colors (green, red, blue), each had one hole (1.3 cm in diameter) drilled through the center so that the ball will fit onto any of the three pegs. A set of 17 x 22 cm cards in a three-ring binder displays the goal states of the individual items that are presented to the participant.

On the TOL-R, the participant must reconfigure three different colored balls on three pegs of different heights. This constrains the number of balls that may be placed on each peg (i.e., the smallest peg can fit only one ball, the middle peg can hold two balls, and the tallest peg can hold all three balls). In addition, only one ball may be moved at a time and the balls must always be held on one of the three pegs during the move sequence.

Procedure

The TOH-R and the TOL-R were administered in a counterbalanced order. The TOH-R was administered in the same manner as described in Phase 1 for the initial pool of items. The 22-item TOH-R represented 7- to 15-move items. The tower-ending and flat-ending items were alternated. Participants were given two practice items, and the subsequent test items were presented in an ascending order of difficulty.

The 30 TOL-R problems were comprised of 10 problems at each of three levels: 4-move, 5-move, and 6-move. Participants were told the number of moves required to solve each problem and were given one point for each problem solved correctly (i.e., conforming to the rules stated earlier). Two practice problems were administered, and then the test items were presented in an ascending order of difficulty. Schnirman et al. (1998) imposed a 2-minute time limit for solution of each item, and although this time limit was not strictly enforced, no participants exceeded this limit on any items.

Results

This independent assessment of the reliability of the 22-item TOH-R yielded an internal consistency of alpha = .77. The total score for TOH-R ranged from 4 to 19 out of a possible 22, and the mean total score was 10.96 ($SD = 4.23$). The mean number correct for the 22-item task was significantly lower than the mean for the same 22 items embedded within the 60-item task in Phase 1, $t(131) = 8.54$, $p < .001$. In addition, there was a significant difference between the performance on tower-ending items ($M = 6.02$, $SD = 2.68$) and flat-ending items ($M = 4.94$, $SD = 2.52$), $t(51) = 3.76$, $p < .0001$.

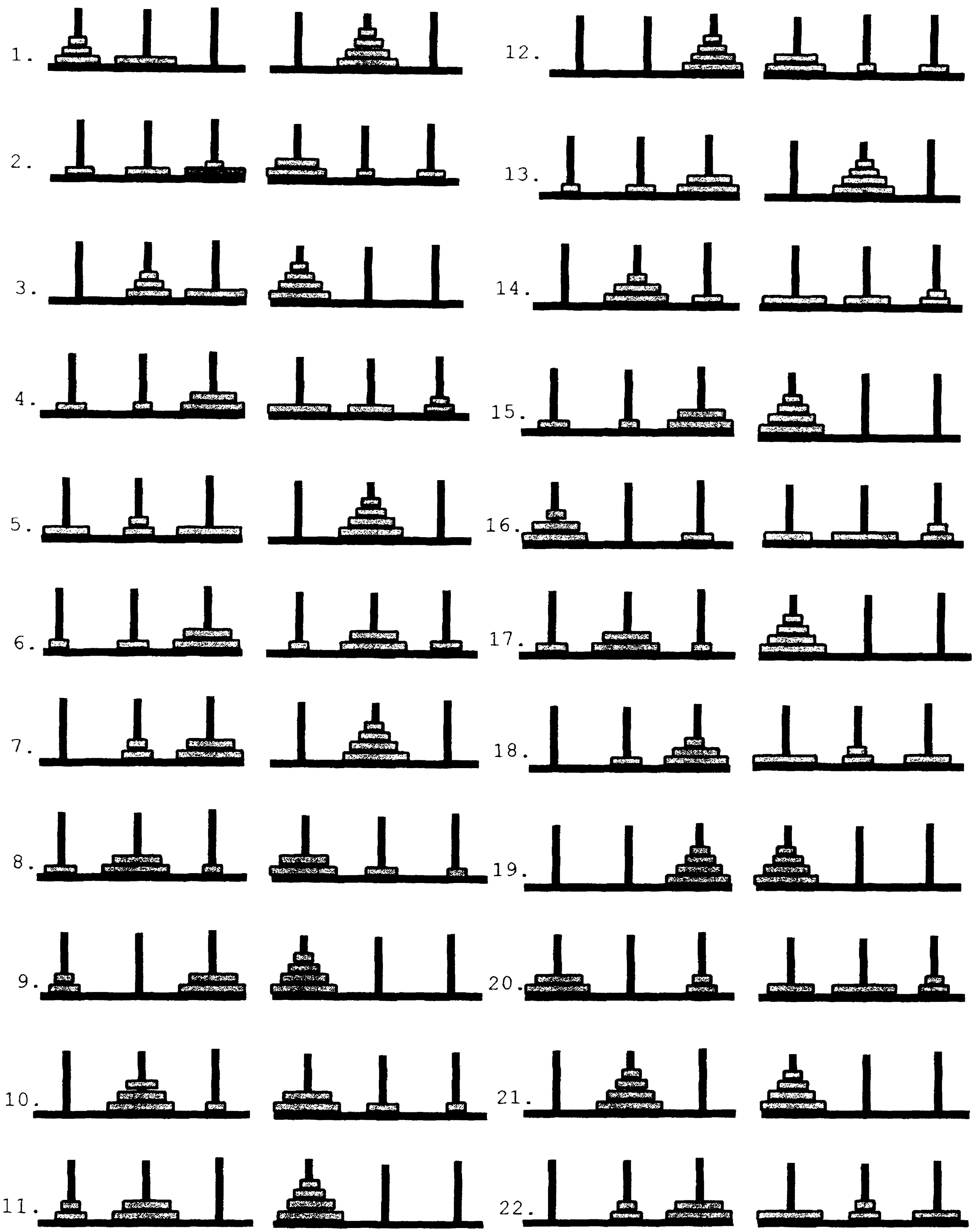

Figure 1. The 22 items comprising the Tower of Hanoi–Revised. For each item, the start state is presented on the left and the end state is presented on the right.

The item-total correlations from this phase of testing were similar to those found in Phase 1 (see Table 1). The range of these correlations was between .05 (an 8-move item) and .50 (a 12-move item).

The internal consistency score for TOL-R replicated the Schnirman et al. (1998) finding (alpha = .70). The total score for TOL-R ranged from 14 to 30 out of a possible 30, and the mean total score was 22.18 (*SD* = 3.99). These performance scores also replicated the findings of Schnirman et al.

The correlation between the TOH-R and TOL-R was $r(48) = .53$, $p < .0001$. To explore the difference between the level of performance on each task, the total scores were converted to percentage correct. There was a significant difference between the mean TOH-R percentage score 49.83 (*SD* = 20.14) and the mean TOL-R percentage score, 73.93 (*SD* = 13.31), $t(49) = 9.86$, $p < .0001$. There was no order effect in testing scores; scores on the TOH-R did not differ if the task was given before ($M = 10.91$, $SD = 4.39$) or after ($M = 10.96$, $SD = .471$) the TOL-R, $t(46) = .036$, $p > .05$.

General Discussion

The purpose of the current study was to redesign the TOH task to increase its reliability and to examine the convergent validity of this task against the TOL. The neuropsychological literature is replete with explicit and implicit references to the two disk-transfer tasks as isomorphic. However, recent research (eg., Welsh et al., 1999) has indicated nonshared variance between the tasks that could be, to some extent, attributed to measurement error resulting from the unreliability of one or both of the tasks. Given that the TOL-R has been redesigned to achieve an acceptable level of internal consistency (Schnirman et al., 1998), the next logical step was to redesign the TOH with the same goal in mind. Only when both tasks have been constructed to achieve an acceptable level of reliability can one test the degree to which performance on the two executive function tasks covary.

In the current redesign of the TOH, two potential explanations for the nonshared variance with the TOL-R were addressed. First, the TOH was reconstructed to be identical to the TOL on key administration features that may have contributed to the lack of overlap between the performances on the two tasks in the past. Second, classical test construction theory was utilized to select, from a larger pool of items, a subset of the most homogeneous items to include in the revised TOH task. This procedure addressed the second explanation for nonshared variance, that is, the lack of reliability of the tasks. The newly designed TOH-R has achieved an acceptable level of internal consistency (alpha = .77), similar to that of the TOL-R (alpha = .70). Therefore, we now have two psychometrically sound tasks that allow us to legitimately explore convergent validity.

The TOH-R and the TOL-R were significantly, albeit moderately, correlated ($r = .53$). This finding was within the range that has been found previously in our laboratory (Humes et al., 1997; Welsh et al., 1999, 2000). The association reflects 72% of the variance that is not shared between the two tasks. Given the redesign of the TOH-R, one can eliminate administration differences as a contributing factor to this nonshared variance. Moreover, given the increased reliability of the TOH-R, one can assume that error variance also contributes less to the nonshared variance. Although previous research in our laboratory yielded TOH-TOL correlations that ranged from .35 to .60, either both or one of these tasks were suspect in terms of reliability. The current significant, albeit moderate, correlation between the two tasks, now suggests that the explanation for the nonshared variance between the TOH-R and TOL-R may involve differential cognitive processes underlying the performance on each task. However, an alternative perspective on the moderate intertask correlation involves the issue of test-retest reliability. Although the test-retest reliability of the TOL-R has been found to be adequate ($r = .70$; Schnirman et.al., 1998), we do not yet know the stability of the newly designed TOH-R. A clearer interpretation of the moderate correlation between the TOH-R and the TOL-R will be possible once the test-retest reliability of the TOH-R is known.

The findings of the Welsh et al. (1999) study indicated different cognitive demands of the two tasks; however, now this question can be addressed using two psychometrically sound measures. Several executive functions, such as planning, working memory, and inhibition, have been proposed to contribute to performance on the TOH and TOL (Goldman-Rakic, 1987; Goel & Grafman, 1995; Pennington & Ozonoff, 1996). However, direct empirical support, in the form of significant correlations between the tower tasks and measures of these other executive functions has been sparse (e.g., Welsh et al., 1999). Clearly, one barrier to such research is the difficulty in finding reliable and valid measures of processes such as planning, working memory, and inhibition. A recent study by Phillips, Wynn, Gilhooly, Della Sala, and Logie (1999) utilized a dual task paradigm and found that performance on the TOL was related to spatial working memory. However, it is important to note that the structure of the TOL used in their study (designed by Ward & Allport, 1997) was substantially different from both the task described by Shallice (1982) and the one used in the current study.

Future work in our laboratory will explore the relative contributions of these executive function processes to the two reliable disk-transfer tasks, the TOH-R and the TOL-R, in an effort to identify the source(s) of the nonshared variance. Currently, we hypothesized that planning, working memory, and inhibition contribute to different degrees to both tasks; however, the two diverge in contribution of a fourth process: rule induction. The structure and the constraints of the TOH are such that a single, rule-based strategy can solve any problem of any level of difficulty; even a partial understanding of the strategy will lead to relatively good performance. This particular rule-based strategy is known as goal recursion, a cyclic algorithm first introduced by Simon (1975). In goal recursion, the entire sequence of correct moves is decomposed into cycles in which progressively smaller subpyramids of disks are moved out of the way in order to move the largest disk (currently not on the goal peg) to its goal position. Although planning, working memory, and inhibition may be necessary to *induce* this strategy, once goal recursion is discovered, it should *reduce* the need for other executive function processes. In contrast, there is no single rule that can be applied to solve TOL problems; one must plan *single* moves for each problem in working memory, while also inhibiting the tendency to make intuitive, but incorrect moves. The hypothesis that the TOH specifically demands rule induction processes is being examined in our laboratory and contrasted with the procedural learning explanation for TOH performance (Gabrieli, 1998; Goldberg et al., 1990).

In summary, an internally consistent version of the TOH, the TOH-R, has been developed, and it correlates moderately with the TOL-R. Performance on the TOH-R by a sample of normal college students indicates that it is a challenging task that elicits substantial variability. If this result is replicated on other normal adult samples, it suggests that the TOH-R may be valuable for exploring normal individual differences in executive functions. However, these task characteristics could pose a problem for use with clinically diagnosed participants, and the applicability of the current task to clinical populations must be tested. Further research must explore the psychometric characteristics of the TOH-R in the context of a more heterogeneous, representative sample as well as specific populations of interest (eg., older, less educated adults; adults diagnosed with attention deficit disorder, etc.). It is entirely possible that, given participants with impaired executive function skills, this same procedure of selecting the most reliable test items might yield a somewhat different composition of TOH task.

This research, in concert with other recent studies, indicates that somewhat different cognitive mechanisms underlie performance on each task. If so, the performance on the TOH-R and TOL-R tasks may prove to be sensitive to differential circuits within the prefrontal cortical system (e.g., Goldman-Rakic, 1998).

References

Brennan, M., Welsh, M. C., & Fisher, C. B. (1997). Aging and executive function skills: An examination of a community-welling older adult population. *Perceptual and Motor Skills, 84*, 1187-1197.

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

Culbertson, W. C., & Zillmer, E. A. (1998). The construct validity of the Tower of London-super (DX) as a measure of the executive functioning of ADHD children. *Assessment, 5*(3), 215-226.

Gabrieli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology, 49*, 87-116.

Glosser, F., & Goodglass, H. (1990). Disorders in executive control functions among aphasic and other brain-damaged patients. *Journal of Clinical and Experimental Neuropsychology, 12*, 485-501.

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

Goldberg, T. E., Saint-Cyr, J. A., & Weinberger, D. R. (1990). Assessment of procedural learning and problem solving in schizophrenic patients by Tower of Hanoi type tasks. *Journal of Neuropsychiatry, 2*, 165-173.

Goldman-Rakic, P. S. (1987). Development of cortical circuitry and cognitive function. *Child Development, 58*, 601-622.

Goldman-Rakic, P. S. (1998). The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. In A. Roberts, T. Robbins, & L. Weiskrantz (Eds.), *The prefrontal cortex: Executive and cognitive functions* (pp. 87-102). New York: Oxford University Press.

Hanes, K. R., Andrewes, D. G., Smith, D. J., & Pantelis, C. (1996). A brief assessment of executive control dysfunction: Discriminant validity and homogeneity of planning, set shift, and fluency measures. *Archives of Clinical Neuropsychology, 11*, 185-191.

Humes, G. E., Welsh, M. C., Retzlaff, P., & Cookson, N. (1997). Towers of Hanoi and London: Reliability of two executive function tasks. *Assessment, 4*, 249-257.

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. *Psychopharmacology, 107*, 394-404.

Levin, H. S., Mendelsohn, D., Lilly, M. J., Fletcher, J. M., Culhane, K. A., Chapman, S. B., Harward, H., Kusnerik, L., Bruce, D., & Eisenberg, H. M. (1994). Tower of London performance in relation to magnetic resonance imaging following closed head injury in children. *Neuropsychology, 8*, 171-179.

Mazzocco, M. M., Hagerman, R. J., Cronister, S. A., & Pennington, B. F. (1992). Problem solving limitations among cytogenetically expressing Fragile X-women. *American Journal of Medical Genetics, 43*, 78-86.

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

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology, 32*, 1081-1105.

Pennington, B. F., Groisser, D., & Welsh, M. C. (1993). Contrasting cognitive deficits in attention deficit hyperactivity disorder versus reading disability. *Developmental Psychology, 29*(3), 511-523.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 37*, 51-87.

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.

Roberts, R. J., & Pennington, B. F. (1996). An interactive framework for examining prefrontal processes. *Developmental Neuropsychology, 12*, 105-126.

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

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. In D. E. Broadbent & L. Weiskrantz (Eds.), *The neuropsychology of cognitive function* (pp. 199-209). London: The Royal Society.

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

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-652.

Stuss, D. T. (1992). Biological and psychological development of executive functions. *Brain and Cognition, 20*, 8-23.

Stuss, D. T., & Benson, F. (1984). Neuropsychological studies of frontal lobes. *Psychological Bulletin, 95*, 3-28.

Ward, G., & Allport, A. (1997). Planning and problem-solving using the 5-disc Tower of London task. *Quarterly Journal of Experimental Psychology, 50*, 49-78.

Welsh, M. C. (1991). Rule-guided behavior and self-monitoring on the Tower of Hanoi disk-transfer task. *Cognitive Development, 62*, 59-67.

Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe function in children: Views from developmental psychology. *Developmental Neuropsychology, 4*, 199-230.

Welsh, M. C., & Pennington, B. F. (1992, February). *A critical evaluation of the current and potential applications of disk-transfer tasks to clinical and experimental neuropsychology*. Paper presented at the annual meeting of INS, San Diego, CA.

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

Welsh, M. C., Pennington, B. F., Ozonoff, S., McCabe, E. R., & Rouse, B. (1990). Neuropsychology of early-treated phenylketonuria: Specific executive function deficits. *Child Development, 61*, 1697-1713.

Welsh, M. C., Revilla, V., Strongin, D., & Kepler, M. (2000). Towers of Hanoi and London: Is the nonshared variance due to differences in task administration? *Perceptual and Motor Skills, 90*, 562-567.

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