The Tower of London\textsuperscript{DX}: A Standardized Approach to Assessing Executive Functioning in Children

William C. Culbertson and Eric A. Zillmer

Neuropsychology Program, Drexel University

In the current study, the Tower of London (Shallice, 1982) was modified to enhance its clinical utility as a measure of childhood executive functioning. The Tower of London-Drexel (TOL\textsuperscript{DX}) was administered to normal control (NC; \( N = 56 \)) and attention-deficit hyperactivity disorder (ADHD; \( N = 99 \)) children (ages 7 to 12) to determine whether age-related changes in performance were evident, to gather normative data, and to evaluate the test-retest reliability and criterion-validity of the measure. The results revealed age-related changes in scores, age-group normative data, an acceptable level of reliability and significant differences in performance of NC and ADHD subjects. Further, discriminant analysis classification rates determined that the TOL\textsuperscript{DX} was sensitive and highly specific to ADHD. Implications and limitations of the study are discussed.

Executive planning represents higher-order cognitive operations necessary for problem-solving and adaptation, which has attracted considerable attention in neuropsychology and related areas. This attention relates to the significance of executive planning in neurobehavioral development (Weinberger, Berman, Gold, & Goldberg, 1994), its integral role in the performance of everyday activities of living (Shallice & Burgess, 1991a), and its association to a number of developmental and acquired neuropsychological disorders (Andreasen et al., 1992; Levin et al., 1994, 1996; Owen et al., 1995; Ozonoff, Pennington, & Rogers, 1991; Pennington & Ozonoff, 1996). The prefrontal lobes in systematic interaction with other cortical and subcortical regions are centrally involved in executive planning. Injury to this network can have profound and disabling behavioral consequences to the developing child (Benton, 1991; Eslinger, 1996).

Executive planning involves the delineation, organization, and integration of behaviors needed to operationalize an intent or achieve a goal. The process of executive planning requires the ability to conceptualize change (anticipate or look ahead), respond objectively, generate and select alternatives, and sustain attention (Lezak, 1995). The assessment of this function, although evolving, is still in its infancy, as is evident in the paucity of measures that are sensitive and specific to planning abilities of children.

Address correspondence to William C. Culbertson, Drexel University, Neuropsychology Program, 32nd and Chestnut Streets, Philadelphia, PA, 19104.
The development of childhood executive measures has been limited by the traditional neuropsychological view that the frontal lobes remain ‘silent’ with regard to function until adolescence or early adulthood (Dennis, 1991). From this perspective, early injury either has minimal immediate consequences (spared function) or does not manifest its impact until the frontal lobes mature. Increasingly, there has been empirical work which suggests that, concomitant with the protracted development of the frontal lobes, are emerging rudimentary prefrontal functions (Levin et al., 1991; Welsh, Pennington, & Groisser, 1991). Accordingly, early damage can have both immediate and late-appearing effects on behavior.

The Tower of London (TOL) was originally developed by Shallice (1982) in an effort to assess executive planning. Reflecting an integration of cognitive and neuropsychological science, Shallice (1982) proposed an information-processing conceptualization of prefrontal functioning. A higher-order, Supervisory Attentional System (SAS) was postulated to modulate the selection, programming, and monitoring of behavior in problem-solving when over-learned, routine, or automatic programs (schema) were insufficient to meet problem-solving demands. Seeking to test the viability of the model, Shallice (1982) sought measures that would assess SAS functions. Borrowing from research in cognitive problem-solving (viz., “artificial intelligence”), the Tower of Hanoi (TOH) was selected as a potential measure due to its novel, anticipatory, and means-end problem-solving demands. However, the TOH was not considered to possess suitable psychometric properties, since it did not allow for the presentation of a variety of different problem configurations of comparable difficulty (Shallice & Burgess, 1991b) and, accordingly, was modified. The modified tower structure (Tower of London) allowed for the presentation of multiple problem configurations of comparable difficulty levels.

In an initial study with the TOL (Shallice, 1982), patients with anterior and posterior cerebral lesions were presented with 12 TOL problems (3-, 4-, and 5-moves). The left frontal patients were found to solve significantly fewer problems within a 60-second time limit and to exhibit greater delay in initiating their first problem-solving move than patients with right anterior and right and left posterior lesions. TOL performance was determined to be independent of intelligence and spatial problem-solving. Subsequent studies (Lange et al., 1992; Mazzocco, Hagerman, Cronister-Silverman, & Pennington, 1992; Morris et al., 1988; Owen, Downes, Sahakian, Polkey, & Robbins, 1990; Owen et al., 1995) have confirmed the value of the TOL as a measure of higher-order problem-solving for normal and clinical groups.

Despite the value of the TOL in the assessment of executive planning, a review of the existing tower systems suggested that several changes would potentially increase the applicability and clinical utility of the measure for childhood populations. A number of tower systems exist that lack standardization in their approach to administering the measure (e.g., Casey, Vauss, Chused, & Swedo, 1994; Saint-Cyr, Taylor, & Lange, 1988). In addition, other systems (Levin et al., 1995; Welsh et al., 1991) require the reapplication of failed problems, which significantly increases the amount of on-task time, a liability when assessing younger children and those with limited attentional capacities (e.g., the traumatic brain injured). Relatedly, required repetitions of failed problems can engender high levels of frustration and distress in children, leading to diminished motivation and cooperation, and disrupted performance. Further, repeated task presentation violates a central task dimension (novel or nonroutine problem-solving demands) necessary for the recruitment of executive planning (Shallice, 1982; Shallice & Burgess, 1991b). Other cognitive operations are likely introduced into the problem-solving process (e.g., procedural learning) when multiple task repetitions are employed, which may alter the sensitivity and/or specificity of the measure to executive planning.

Increasingly, neuropsychological measures are being computerized, including TOL structures (Andreasen et al., 1992; Leon-Carrion et al., 1991; Owen et al., 1995; Saint-Cyr et al.,...
Computerized administration of the TOL, however, may not be that effective for the following reasons. First, the utilization of computerized measures can tax the limited attentional resources of brain-impaired individuals, particularly those who are very concrete in their problem-solving orientation. Similarly, young children whose abstract abilities have not fully emerged, problem-solve more efficiently when manipulative materials are employed. Second, the three-dimensional, manipulative nature of the TOL allows the examiner to observe qualitative aspects of the subject’s executive planning that are not afforded by a computerized measure. For example, the manner in which a rule violation is committed can be very informative. Namely, subjects may violate the TOL problem-solving rules without awareness, whereas others may do so with intent, as evidenced in efforts to conceal their rule breaking (viz., waiting for the examiner to look away, shielding bead moves from the examiner’s view, etc.). Finally, the TOL requires examiner-subject interaction and thus provides information related to executive planning within a social context.

The current study proposes a number of modifications of the TOL to enhance its clinical and empirical utility as a measure of executive functioning. To examine these changes, the modified Tower of London-Drexel (TOL-DX; Culbertson & Zillmer, 1995) was administered to two groups, a sample of normal children and a group of clinic-referred children, to determine whether age-related changes in performance were evident across scoring variables, to gather preliminary normative data, and to evaluate the reliability and validity of the measure. It was hypothesized that normal and clinic-referred children would exhibit age-related changes in TOL-DX performance. Further, it was predicted that the measure would assess executive planning in a reliable and valid manner. Specifically, the TOL-DX would be found to demonstrate acceptable test-retest reliability and criterion-validity to allow for clinical use.

**METHOD**

**Subjects**

Two cross-sectional samples were employed in the initial development and standardization of the TOL-DX. The first sample consisted of normal children (NC; \(N = 56\)) between the ages of 7 and 12 years (\(M = 120.7\) months, \(SD = 20.3\) months) who were drawn from a suburban, middle-class community. Teachers in the school district were asked to nominate children who exhibited at least average academic performance in the classroom. Report cards were reviewed to insure that nominated students demonstrated a history of at least average academic achievement and did not exhibit behavioral or work study problems. Further, group achievement scores (Iowa Test of Basic Skills, Riverside Publishing, 1993) were available for 44 of the children. The achievement scores were examined to identify children with possible skill deficits (i.e., below-average achievement score). Children with a history of academic or behavioral difficulties were excluded from the study. Further, students who received special educational or remedial assistance, psychotherapeutic services, and/or had been retained in a grade, were not entered into the study. Of the 70 students initially referred, 11 were excluded from, or decided not to participate in, the study.

The sample was composed primarily of Caucasian children (96%) with an equal representation of males and females, \(n = 28\), respectively. The children were divided into two groups based upon age. Specifically, children ages 7 through 9 comprised the first group \((n = 28)\), whereas subjects 10 through 12 years of age defined the second group \((n = 28)\). The two groups did not differ in terms of their achievement performance (composite standard score) on the Iowa Test of Basic Skills, \(t(42) = 1.2, p < .2\). Further, when the performance of the children was contrasted by gender, male and female subjects did not significantly differ in their achievement, \(t(42) = .4, p = .7\). Consent to be involved in the study was gained from
each child and the parent of each child. Each subject was tested individually by a licensed psychologist within a quiet office located in the child’s school.

The second sample was comprised of children exhibiting an attention-deficit hyperactivity disorder (ADHD). This clinical group was selected due to its posited deficits in executive self-regulation, particularly as related to higher-order planning and organization (Shue & Douglas, 1991). Insofar as the TOL DX was developed to assess executive planning, the performance of ADHD was expected to differ from that of the NC children. The ADHD sample was comprised of children between 7 and 12 years of age (N = 99). The children were drawn from consecutive referrals to a clinic specializing in the neuropsychological evaluation and treatment of ADHD. The majority of the children (n = 72) comprising the sample were male, a finding consistent with previous findings of the proportion of male to female children who are referred for clinic services (Barkley, 1997; Culbertson & Krull, 1996). The children did not present histories of mental retardation, severe psychiatric disturbance, or neurological injury/disorder. However, all the children met the criteria for the diagnosis of ADHD (DSM III-R, American Psychiatric Association, 1987) as determined by structured parent interview, teacher and parent rating scales, and objective neuropsychological testing by a licensed psychologist. In many cases, the children had been previously diagnosed by other psychologists or medical professionals.

Comorbid disorders were exhibited by a number (n = 46) of the ADHD children. These comorbid conditions included the following: oppositional defiant/conduct disorders (23%), anxiety disorders (9%), depressive disorders (2%), adjustment disorders (3%), and learning disabilities (9%). Table 1 presents the demographic characteristics, selected intellectual scores (Wechsler Intelligence Scale for Children III; Wechsler, 1991), and behavioral rating scores (Child Behavior Checklist for Ages 4–18; Achenbach, 1991; Revised Conners Parent Rating Scale; Goyette, Conners, & Ulrich, 1978) for the ADHD sample stratified by age groups. Clearly, the ADHD group exhibited elevated levels of inattention, impulsivity, and overactivity.

**Procedure**

The TOL DX test materials includes two identical tower structures, one for the child and one for the examiner. Each structure consists of three wooden pegs of descending heights.
mounted on a block base. Three beads (red, green, and blue) are placed on the pegs in a prescribed start position (see Figure 1). The subjects are informed that they are to replicate, on their tower structure, the different problem configurations that are to be presented on the examiner’s tower structure. Further, the subjects are told to solve the presented problems in as few moves as possible, while adhering to two rules. The first rule requires the child to move the beads from the pegs one at a time, and the second prohibits efforts to place more beads on a peg than it will accommodate.

The TOL DX is administered in accordance with precise standardized instructions (Appendix A) presented in a detailed manual that also provides guidelines for scoring and interpretation. The instructions employ simple, clear, and specific language to facilitate comprehension across the ages. The directions can be reworded to enable children with specific needs or disabilities to understand the demands of the measure. Time of administration is approximately 20 minutes.

As previously discussed, multiple trials for failed items is not considered an appropriate assessment strategy. Accordingly, the child is allowed only one trial per TOL DX test item. Further, 2- to 5-move test items have traditionally been employed to assess executive functioning (Kirkorian, Bartok, & Gay, 1994; Shallice, 1982). In an effort to enhance the sensitivity of the measure, 6- and 7-move problems are introduced to allow for a “finer grained” assessment of executive planning (see Appendix B). A total of 15 test problems are sequen-
tially presented to the child, each being assembled on the examiner’s tower structure. The
test problems are presented in Appendix B.

Scoring

A variety of scaling metrics have been employed in scoring the performance of subjects
with tower structures (Levin et al., 1996; Morris, Ahmed, Syed, & Toone, 1993; Pennington,
Groisser, & Welsh, 1993; Welsh, 1991). To potentially maximize the sensitivity of the TOL\textsuperscript{DX}
to executive problem-solving performance, an interval scale was selected based on the num-
ber of moves to solution. Specifically, the number of moves executed by the subject minus
the minimum number of moves to solution was employed. The resulting score dimension
allows for the representation and differentiation of a wide range of individual differences in
executive problem-solving that may exist across childhood populations.

Six scores are computed to include move score, time violations, rule violations, initiation
time, solution time, and total problem-solving time. The move score involves a summation
of bead moves exceeding the optimal number of moves necessary to solve test problems.
Any score exceeding 20 moves is automatically assigned a move score of 20 to avoid the
inflation of the total move score by excessive moves on a limited number of items, to limit
untoward frustration for the children and to be consistent with the move per item cut-off
utilized by previous tower systems (Pennington et al., 1993; Welsh et al., 1991). For example,
if a child executes 24 moves on a 7-move problem, the score would be calculated as follows:
\[20 - 7 = 13\]. Additionally, a score of 20 is assigned for any item not solved within 2 minutes.

The child is allowed 2 minutes per test item, but is assessed a time violation at 1 minute
if the test problem is not solved. The 1-minute criteria provides a measure of the efficiency
of executive planning and problem solving. Rule violations pertain to moving two beads off
the pegs at the same time or efforts to place more beads on a peg than it will hold. The total
number of rule and time violations are calculated.

Initiation time is defined as the time from the presentation of a test problem by the exam-
iner to the initiation of the first problem-solving move (lifting of a bead from a rod). Solution
time entails the time from initiation of the first move to the completion or discontinuation
of problem-solving. The interval from the presentation of a test problem to its solution or
discontinuation of performance by the child defines total problem-solving time (i.e., problem-
solving time = initiation time + solution time).

RESULTS

Gender effect was examined to determine whether significant TOL\textsuperscript{DX} performance differ-
ences exist. For the nonclinical group, younger male and female children, when grouped by
age (7 to 9 years) demonstrated comparable TOL\textsuperscript{DX} move, time and rule violation perfor-
mancess. \[t(26) = .5, p = .6, t(26) = -.3, p = .8, \text{ and } t(26) = -.8, p = .4,\] respectively.
For the older children (10 to 12 years), a gender difference was not evident for move score,
\[t(26) = .5, p = .6, \text{ or time violation score, } t(26) = -1.6, p = .1.\] Older female children
showed a tendency to commit more rule violations \(M = 1.1, SD = 1.8\) than their male
cohorts \(M = .1, SD = .3,\) \[t(26) = -2.0, p < .06,\] although their performance fell short
of significance. Accordingly, male and female scores were combined for analysis.

Due to the small number of ADHD females when split into the two age groups, gender
by age group contrasts were not computed. Rather, the age groups were collapsed and the
performance of the ADHD females \(n = 27\) was contrasted to that of the ADHD males
\(n = 72.\) The comparisons failed to reveal any significant difference in gender performance
for move score, \( t = -0.5, p = 0.6 \), time violations \( t = -0.3, p = 0.7 \) or rule violations, \( t = 1.2, p = 0.2 \).

The TOL\textsuperscript{DX} scores for the NC and ADHD children by individual age levels were contrasted. A one-way analysis of variance (ANOVA) revealed a significant decrease in TOL\textsuperscript{DX} total move score by increasing individual age levels for the NC and ADHD children, \( F(5, 50) = 3.1, p < 0.02 \) and \( F(5, 93) = 10.6, p < 0.0001 \), respectively. Age-related score differences (see Figure 2) are evident with older NC and ADHD children achieving lower raw move scores (greater proficiency) than their younger cohorts.

Age-related performance for the total time and rule violation scores (see Figure 3) was also investigated. The mean total time violation scores were found to decrease significantly as a consequence of increasing individual age levels, \( F(5, 50) = 3.5, p < 0.009 \) for NC children. In contrast, the NC children did not differ in the number of rule violations committed as a function of age level, \( F(5, 50) = 0.9, p = 0.5 \). Across the age levels, NC children generally committed very few rule violations. Overall, older NC children were more likely than younger children to solve test problems within the 1-minute time limit, with both groups performing in a rule-governed manner.

The ADHD children also exhibited changing time violation performance with increasing age, \( F(5, 93) = 11.4, p < 0.0002 \). However, unlike the NC children, they demonstrated a
significant reduction in rule violations by age, $F(5, 93) = 8.1, p < .0002$. Thus, with increasing age, both time and rule violation scores decreased for ADHD children.

With regard to time variables, only total initiation, solution and problem-solving times were available for NC children (see Figure 4). Their temporal performance was subjected to analysis (one-way ANOVA), which revealed that the NC children differed, by age, in total solution time, $F(5, 50) = 4.7, p < .002$ and total problem-solving time, $F(5, 50) = 3.4, p < .01$ time, but not initiation time, $F(5, 50) = .4, p = .8$). Thus, younger children needed more time to solve test items, which contributed to a longer total problem-solving time as compared to older cohorts.

**Standard Scores**

Table 2 presents the means and standard deviations (raw scores) of the TOL\(\text{DX}\) for the NC and ADHD children. For normative utilization, the total move and time raw scores were transformed to scores and then standardized with a mean of 100 and a standard deviation of 15. The total time violation and rule violation raw scores were also transformed to scaled scores, but with a mean of 10 and a standard deviation of 3. The score transformations were computed by age groups (7 to 9 and 10 to 12 years). Conversion tables were developed to allow for the quick transformation of the raw score variables to standard scale equivalents.\(^1\)

\(^1\)The conversion tables of raw scores to standard score equivalents can be obtained from the first author.
FIGURE 4. Mean TOL\textsuperscript{DX} time score by age level for normal control (NC) children ($N = 56$).

TABLE 2

<table>
<thead>
<tr>
<th></th>
<th>NC ($n = 28$)</th>
<th>ADHD ($n = 56$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Total Move Score</td>
<td>61.3</td>
<td>20.9</td>
</tr>
<tr>
<td>Total Time Violations</td>
<td>.6</td>
<td>.9</td>
</tr>
<tr>
<td>Total Rule Violations</td>
<td>59.3</td>
<td>27.7</td>
</tr>
<tr>
<td>Total Initiation Time (sec)</td>
<td>59.3</td>
<td>27.7</td>
</tr>
<tr>
<td>Total Solution Time (sec)</td>
<td>450.1</td>
<td>172.9</td>
</tr>
<tr>
<td>Total Problem Solving Time (sec)</td>
<td>509.5</td>
<td>186.4</td>
</tr>
<tr>
<td>Total Move Score</td>
<td>46.5</td>
<td>16.9</td>
</tr>
<tr>
<td>Total Time Violations</td>
<td>1.0</td>
<td>1.3</td>
</tr>
<tr>
<td>Total Rule Violations</td>
<td>.6</td>
<td>1.4</td>
</tr>
<tr>
<td>Total Initiation Time (sec)</td>
<td>68.2</td>
<td>53.8</td>
</tr>
<tr>
<td>Total Solution Time (sec)</td>
<td>318.0</td>
<td>108.8</td>
</tr>
<tr>
<td>Total Problem Solving Time (sec)</td>
<td>386.1</td>
<td>120.8</td>
</tr>
</tbody>
</table>

Note. NC = normal controls; ADHD = attention-deficit hyperactivity disorder.
*Raw scores.
TABLE 3

TOL\textsuperscript{DX} Test-Retest Reliabilities for ADHD Children (N = 30)

<table>
<thead>
<tr>
<th>Age Group</th>
<th>TOL\textsuperscript{DX} Scores</th>
<th>Time 1</th>
<th>Time 2</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>7–10 Total Test Score</td>
<td>85.9</td>
<td>30.0</td>
<td>69.1</td>
<td>23.7</td>
</tr>
<tr>
<td></td>
<td>Total Time Violations</td>
<td>3.5</td>
<td>2.9</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>Total Rule Violations</td>
<td>3.3</td>
<td>3.3</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Note. ADHD = attention-deficit hyperactivity disorder.
*p < .05. **p < .005.

It should be noted that increasing standard scaled scores correspond to lower raw move counts (i.e., greater proficiency), fewer rule and time violation scores, and quicker initiation, solution and problem-solving time scores.

Reliability and Validity of the TOL\textsuperscript{DX}

An analysis of the temporal stability of the TOL\textsuperscript{DX} scores (total move, time violation, and rule violation scores) was undertaken with a subgroup of 7- to 10-year-old children (n = 30), who were exhibiting an attention-deficit/hyperactivity disorder (ADHD). These children were selected to provide a stringent test of the stability of the TOL\textsuperscript{DX}, since ADHD children are prone to exhibit relatively high levels of performance variability (Barkley, 1990). The ADHD children were assessed on two occasions in a standardized manner with the temporal interval between assessment averaging 16.3 days (SD = 8.9; range 7 to 41 days).

Table 3 presents the test-retest coefficients. The reliability coefficient of the TOL\textsuperscript{DX} total move score is within the moderate to high range (r = .81, p < .005) and indicates that the move score is relatively stable across time. A decrease of approximately 17 points in total move score can be expected between testings of equivalent temporal intervals. The decrease in score likely reflects the effects of learning and practice.

The temporal reliability coefficient of the TOL\textsuperscript{DX} total time violation score is at a moderate level (r = .79, p < .005). As of the second testing, the children averaged approximately 1.5 fewer total time violations as compared to their performance during the initial testing (see Table 2). The moderate level of stability indicates that time violation performance remains relatively constant across time intervals. In contrast, the stability of the TOL\textsuperscript{DX} rule violation score is low (r = .42, p < .05). Table 4 reveals a reduction in mean rule violation scores, at Time 2, of approximately 2 violations. Thus, the children tended to perform in a more

TABLE 4

Classification Rates of ADHD and NC Children

<table>
<thead>
<tr>
<th>Predicted Group</th>
<th>Actual Group</th>
<th>N</th>
<th>ADHD</th>
<th>NC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHD</td>
<td>99</td>
<td>63 (64%)</td>
<td>36 (36%)</td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>56</td>
<td>11 (20%)</td>
<td>45 (80%)</td>
<td></td>
</tr>
</tbody>
</table>

Note. ADHD = attention-deficit hyperactivity disorder and NC = normal controls.
Positive predictive power = .85; negative predictive power = .56 and overall classification rate = .70.
rule-governed manner at the second time of testing. However, the low reliability coefficient signals that the change in performance across time was not strongly related to the child’s initial level of performance. That is, the majority of children committed few rule violations at Time 2, regardless of their initial level of performance. Thus, the total rule violation score appears particularly susceptible to change over time.

To investigate the criterion-related validity of the TOL\textsuperscript{DX} performances of the younger (7 to 9 years) and older (10 to 12 years) NC and ADHD children were compared. It was predicted that the ADHD children would demonstrate poorer TOL\textsuperscript{DX} performance than the NC children due to their deficits in regulatory control (Shue & Douglas, 1992). The planning and problem-solving performance (total move score) of ADHD children, 7 to 9 years of age, was found to be significantly poorer than that of the comparable NC age group, \( t(82) = 5.7, p < .001 \), with the ADHD group executing a significantly greater number of extra moves in solving TOL\textsuperscript{DX} test items. Likewise, the older ADHD children employed a significantly greater number of moves in performing the TOL\textsuperscript{DX} test items as compared to their normal cohorts, \( t(69) = 3.9, p < .001 \). Further, the younger ADHD group (7 to 9 years) differed significantly from the NC subgroup with regard to time violations, \( t(82), p < .007 \), while the older ADHD group (10 and 12 years) approached significance in committing a greater number of time violations, \( t(69), p < .06 \) than their NC age cohorts. Finally, both ADHD groups exhibited more rule violations as contrasted to the rule-governed performance of the two like-age control groups, \( t(82) = 5.7, p < .001 \), and \( t(69) = 2.2, p \geq .03 \), respectively.

The diagnostic utility of the TOL\textsuperscript{DX} was investigated by calculating the sensitivity, specificity, and positive and negative predictive power rates of the measure. The performance of ADHD and NC children previously described was subjected to discriminant analysis. Two score variables (total move and rule violation score) were entered for analysis. Table 4 presents the resulting classification rates. The derived discriminant function produced an overall classification rate of .70. The sensitivity and specificity rates were found to be .64 and .80, respectively. These identification rates indicate that approximately 6 out of 10 ADHD children and 8 out of 10 NC children were accurately identified by their respective TOL\textsuperscript{DX} score performance. Moreover, when the TOL\textsuperscript{DX} discriminant function was employed to predict the presence or absence of the disorder (positive and negative predictive power), poor performance was found to be highly predictive of ADHD (.85) while efficient performance was predictive of the absence of the disorder (.54), but to a lesser degree. Overall, the classification rates are acceptable for individual diagnostic decisions, particularly if the TOL\textsuperscript{DX} is utilized in conjunction with other measures of neuropsychological functioning.

**DISCUSSION**

The current cross-sectional study sought to standardize the instruction and scoring, and determine the clinical utility of the recently developed TOL\textsuperscript{DX}. The TOL\textsuperscript{DX} performance of normal and ADHD children was employed to provide preliminary normative data and to investigate the reliability and validity of the measure. A number of significant findings were elucidated.

The TOL\textsuperscript{DX} scores (move, time violation, and rule violation) show an age-related difference for both normal and ADHD children characterized by increasing executive planning and problem-solving efficiency as a function of increasing age. This pattern is consistent with previous studies suggesting that executive planning and problem-solving abilities are more advanced for older as contrasted to younger children (Levin et al., 1991; Welsh et al., 1991).

Age-related differences were not evident for rule violations with normal children, but were
clearly observed for ADHD children. The difference suggests that both younger and older normal children are able to meet the problem-solving demands of the TOL\textsuperscript{DX} while adhering to ‘rule constraints’ governing performance. In fact, normal children across the age levels exhibited a very low rate of rule violations, which likely accounts for the failure to find a significant age-related change. In contrast, ADHD children show a significant decrease in rule violations with age, indicating that older children are more proficient than their younger cohorts in guiding planning and problem-solving via verbal mediation, that is, internalized rules. Further, when the rule governed performance of young ADHD and normal children was compared, the former exhibited a significantly greater number of violations. These differences attenuated with age, although a tendency for more rule violations still existed for older ADHD children.

When the TOL\textsuperscript{DX} performance of the normal and clinic children were contrasted, both age groups of normal children exhibited superior performance across scoring variables. These significant differences are consistent with previous investigations (Grodzinsky & Diamond, 1992; Pennington & Ozonoff, 1996) that have determined that ADHD children exhibit major deficits in regulatory control across childhood age levels, that is, the deficit pattern is relatively invariant.

The time variables also show age-related changes with younger normal children (7 to 9 years) exhibiting slower solution and total problem-solving times than older (10 to 12 years) children. Although motor speed and dexterity effects solution and ultimately total problem-solving times, the significantly greater number of moves employed by the younger children to solve TOL\textsuperscript{DX} problems implicates greater executive efficiency as the primary determinant of improved time performance for older children. The finding of comparable initiation times suggests that both younger and older children plan at similar rates, but the efficiency of planning is greater for older children, as evidenced in their lower move and violation scores. Thus, executive planning performance of older children is characterized by greater accuracy and speed of performance.

The test-retest reliability for ADHD children was found to demonstrate suitable stability over a 2-week period. As previously noted, the ADHD children were selected as a stringent test of the stability of the measure due to their high rates of variability in performance (Barkley, 1990). Additional reliability studies are warranted with normal children to determine if comparable or improved levels of stability are evident. The very specific and detailed nature of the administration and scoring of the TOL\textsuperscript{DX} should help reduce examiner differences in administration and scoring that can potentially attenuate the reliability of the measure.

The criterion-validity of the TOL\textsuperscript{DX} was supported by the significant difference in performance of ADHD children, both younger and older, from comparable normal children across scoring variables. The ADHD children demonstrated significantly poorer and inefficient executive planning and problem-solving. Poorly planned and disorganized behaviors are documented failings of ADHD children (Shue & Douglas, 1992; Zentall, Harper, & Stormont-Spurgin, 1993), and a major detriment to their success, both academically and socially.

The TOL\textsuperscript{DX} demonstrated appropriate classification rates to support its clinical utilization. The measure is sensitive and highly specific to ADHD, with a poor TOL\textsuperscript{DX} performance predicting the presence of the disorder 9 out of 10 times. The sensitivity and negative predictive rates were potentially attenuated by the heterogeneity of ADHD subtypes and comorbid symptom patterns represented in the ADHD sample. The classification rate may be even higher for specific subtypes or comorbid patterns. For example, ADHD children, who demonstrate inattention, impulsivity, and overactivity may show greater impairment on the TOL\textsuperscript{DX} than those ADHD children who exhibit only inattention. Frontal-striatal dysfunction has been implicated in disinhibition, self-regulatory deficits, and impaired attention (Teeter & Semrud-Clikeman, 1995), characteristics that are exhibited by ADHD children of the former type;
whereas brain stem and posterior regions, such as the parietal lobes disruption has been considered to play a causative role in the latter (Goodyear & Hynd, 1992; Posner, 1992). Accordingly, poorer performance would be predicted in the inattentive, impulsive, and over-activity ADHD child due to the putative role of the frontal-striatal system in executive planning and problem-solving (Pennington & Ozonoff, 1996; Pennington, Bennetto, McAleer, & Roberts, 1996; Roberts & Pennington, 1996).

The finding of comparable TOL\textsuperscript{DX} performance for female and male children is consistent with previous studies of tower structures (Krikorian et al., 1994; Welsh et al., 1991). However, investigations with larger samples are needed to insure that gender effects are, in fact, not evident. Relatedly, demographic variables, such as socioeconomic status, race, and geographic locale, were not controlled in either the normal or clinical samples. Replication of the current study’s findings with larger samples in which the aforementioned variables are controlled is needed. Finally, the TOL\textsuperscript{DX} warrants cross-validation with independent samples of ADHD and normal children.

In conclusion, preliminary reliability and validity results are supportive of the TOL\textsuperscript{DX} as a measure of executive planning and problem-solving. Detailed instructions for administration and scoring are available. Several scoring dimensions reflecting different, though related, aspects of executive planning and problem-solving are presented. The measure is well-tolerated by younger populations due to its brevity, novelty and game-like appeal.

REFERENCES


APPENDIX A
Administration Instructions

DEMONSTRATION PROBLEMS (MOVES)

Prior to each trial place the subject’s beads in the start configuration (see Figure 1), then reconfigure the examiner’s board into the next Demonstration, Practice or Test Problem. Place the beads on the designated pegs, one at a time, moving from your left to right.

Point to the respective tower boards which initially are both in the start configuration and say: SEE THESE TWO BOARDS? THEY ARE BOTH ALIKE. THIS BOARD WILL BE THE ONE YOU’LL BE USING AND THIS ONE WILL BE MINE.

Point to the beads on the examiner’s tower board and say: I AM GOING TO PUT MY BEADS ON THE PEGS IN DIFFERENT WAYS, IN DIFFERENT ARRANGEMENTS AND SEE IF YOU CAN MAKE WHAT I MAKE IN AS FEW AS MOVES AS POSSIBLE (emphasize “in as few as moves as possible”). WATCH ME.

Remove all the beads from the examiner’s board and set up the Demonstration Problem. First place the green bead on Peg 1, then the blue bead on Peg 2, and finally, the red bead on Peg 3. Then say: SEE IF YOU CAN MAKE ONE JUST LIKE MINE IN AS FEW MOVES AS POSSIBLE.

If the subject completes the Demonstration Problem correctly (moving the red bead to Peg 3, a 1-move solution), say: WELL DONE.

If the subject fails the Demonstration Problem (extra moves), place his/her beads back into the start configuration and demonstrate the solution. Prompt the child to try again. Continue to demonstrate the solution until the subject understands what is required. Discontinue if the subject is unable to solve the Demonstration Problem. DEMONSTRATION PROBLEMS (RULES)

Move the subject’s beads into start configuration and say: NOW, THERE ARE TWO RULES YOU HAVE TO FOLLOW WHEN YOU ARE ARRANGING THE BEADS TO MAKE THEM JUST LIKE THE ONES ON MY PEGS. THE FIRST RULE IS THAT YOU HAVE TO MOVE YOUR BEADS ONE AT A TIME. YOU CANNOT HAVE TWO BEADS OFF THE PEGS AT THE SAME TIME. THAT MEANS, YOU CANNOT HAVE TWO BEADS “IN THE AIR” AT THE SAME TIME OR ONE BEAD “IN THE AIR” AND ANOTHER ON THE TABLE OR BOARD.

Using the subject’s board, demonstrate moving the beads, one at a time, in different directions while saying: NOTICE HOW I ALWAYS PUT THE BEAD BACK ON A PEG BEFORE MOVING THE NEXT ONE. NO MATTER WHAT DIRECTION YOU MOVE A BEAD, ALWAYS PLACE IT ON A PEG BEFORE SLIDING ANOTHER ONE OFF.

Demonstrate moving a bead from peg 1 to peg 3, and then back from peg 3 to peg 1, while saying: THE RULE DOESN’T CHANGE WHEN YOU JUMP OR MOVE A BEAD OVER A PEG. ALWAYS PLACE THE BEAD ON A PEG BEFORE LIFTING THE NEXT ONE OFF ITS PEG.

Demonstrate variations of having two beads off the pegs at the same time (i.e., simulta-
neously removing the red and green beads from Peg 1, simultaneously removing the red bead from Peg 1 and the blue bead from Peg 2, and removing the blue bead from Peg 2 and placing it on the table and then moving the red bead from Peg 1 to Peg 2) while saying:

**HERE ARE SOME EXAMPLES OF BREAKING THE RULE.**

After several demonstrations, ask the subject: **DO YOU UNDERSTAND THE RULE?**

If the subject expresses a lack of understanding, repeat the demonstrations before proceeding. Once the subject comprehends the rule, return the subject’s beads to the start position, and say: **THE SECOND RULE IS THAT YOU ARE NOT TO PLACE MORE BEADS ON A PEG THAN IT WILL HOLD.**

Move a second bead to Peg 2 on the subject’s board. Then hold a third bead over Peg 2 while saying: **THE SECOND PEG CAN ONLY HOLD TWO BEADS, IT CANNOT HOLD A THIRD BEAD.**

Move a bead to Peg 3 on the subject’s board. Then hold a second bead over Peg 3 while saying: **THE THIRD PEG CAN ONLY HOLD ONE BEAD, IT CANNOT HOLD A SECOND BEAD. DO YOU UNDERSTAND THE RULE?**

If the subject indicates yes, proceed to the Practice Problems. If not, re-explain and re-demonstrate. **PRACTICE PROBLEMS 1–2 (2 MOVES).**

Place the subject’s beads in the start configuration and set up Practice Problem 1 on the examiner’s board. Point to the examiner’s board and say: **NOW MAKE ONE LIKE THIS ON YOUR BOARD IN AS FEW MOVES AS YOU CAN.**

If the subject violates either Rule 1 or 2, immediately stop the subject and return the beads to their previous pegs while saying: **YOU BROKE ONE OF THE RULES, REMEMBER YOU CANNOT HAVE MORE THAN ONE BEAD OFF A PEG AT THE SAME TIME (or YOU CANNOT PUT MORE BEADS ON A PEG THAN IT WILL HOLD).**

Then say: **GO AHEAD, FINISH MAKING YOUR BEADS LOOK LIKE MINE WITHOUT BREAKING THE RULES(S).**

If the subject fails the Practice Problem by making extra moves, return the subject’s beads to the start configuration while saying: **THAT WAS A NICE TRY, BUT YOU MADE EXTRA MOVES. WATCH ME DO IT.**

Demonstrate the correct number of moves using the subject’s board. Then, return the subject’s board to the start configuration and say: **NOW YOU TRY IT AGAIN AND MAKE YOURS LIKE MINE IN AS FEW MOVES AS POSSIBLE.**

If the subject fails the Practice Problem on the second attempt, continue to demonstrate until the subject can solve the problem correctly before proceeding to Practice Problem 2. When the subject correctly solves Practice Problem 1, proceed immediately to Practice Problem 2. Discontinue testing if the subject cannot successfully solve Practice Problem 1 or 2 with repeated trials and demonstrations. **TEST PROBLEMS 1–15**

Place the subject’s beads into start configuration and say: **NOW I AM GOING TO SET UP MORE BEAD PATTERNS AND SEE IF YOU CAN MAKE THEM ON YOUR BOARD IN AS FEW MOVES AS POSSIBLE. YOU MAY FIND THAT SOME OF THE PATTERNS ARE DIFFICULT, BUT DO THE BEST YOU CAN. EACH PATTERN CAN BE SOLVED.**

Administer the remaining test problems. Following the set up of each test problem on the examiner’s board, point to the model and say: **NOW MAKE ONE LIKE THIS.**
Appendix B

FIGURE B1. TOL.png test configurations.