LIFESPAN PERFORMANCE ON A COMPUTERIZED TOWER OF LONDON-
DREXEL: FROM 16 TO 80 YEARS OF AGE

by

KEVIN JAMES MOORE

B.A., Northern Arizona University, 2013

A thesis submitted to the Graduate Faculty of the
University of Colorado Colorado Springs
in partial fulfillment of the
requirements for the degree of
Master of Arts
Department of Psychology
2017
This thesis for the Master of Arts degree by

Kevin James Moore

has been approved for the

Department of Psychology

by

Hasker P. Davis, Chair

Leilani Feliciano

Frederick Coolidge

Date July 12, 2017
Executive function is a set of higher-order cognitive processes that are comprised of the skills and abilities necessary for planned and deliberate, goal-directed behaviors. One of the quintessential examples of executive functioning is that of planning. The Tower of London-Drexel Edition (TOL\textsubscript{DX}) is one of the most researched commercially available tests of planning ability. The TOL\textsubscript{DX} is a tower transfer task designed as a manual test utilizing stimulus boards and a trained experimenter. The present study examined the utility of a computerized version of the TOL\textsubscript{DX}. A total of 930 participants ranging from 16 to 80 years of age completed the computerized TOL\textsubscript{DX}. Age had a strong effect on all measures of performance; excess moves ($F(4, 643) = 7.18, p < .001$), total initial time ($F(4, 645) = 7.21, p < .001$), total execution time ($F(4, 633) = 43.19, p < .001$), and total completion time ($F(4, 634) = 41.6, p < .001$), with older participants performing worse than younger participants. A series of regressions and correlations indicated factors that may predict TOL\textsubscript{DX} performance and found age, fluid intelligence, visuospatial ability, and working memory had weak relationships with TOL\textsubscript{DX} performance. When compared to the original data provided by the creators of the manual TOL\textsubscript{DX}, differences in performance scores indicated that the computerized TOL\textsubscript{DX} may measure different constructs than the manual version. Despite the differences, the
computerized TOL\textsubscript{DX} shows promise as a measure of planning ability and the lifespan norms provides the basis for the use of a computerized version of the TOL\textsubscript{DX}. 
# TABLE OF CONTENTS

## CHAPTER

I. **INTRODUCTION** .................................................................................................1
   - Origin and Development of Tower of London (TOL)...........................1
   - TOL-Drexel 2nd Edition (TOLdx).........................................................4

II. **REVIEW OF THE LITERATURE** ...............................................................7
   - TOL Background .....................................................................................8
   - Planning Ability ......................................................................................10

III. **METHOD** ......................................................................................................15
   - Participants ..............................................................................................15
   - Instruments and Design .........................................................................16

IV. **RESULTS** ......................................................................................................21
   - ANOVA and Post Hoc ..........................................................................21
   - Regression Analyses ...........................................................................24
   - Welch’s T-Tests ....................................................................................27

V. **DISCUSSION** ..................................................................................................30
   - Discussion of Results ..........................................................................31
   - Limitations and Future Considerations ..............................................34

REFERENCES ..........................................................................................................37

APPENDIX A: Correlation Analyses for Outcome and Predictor Variables........42

APPENDIX B: IRB Approval ..................................................................................43
LIST OF TABLES

TABLE

1. Demographics of Sample by Age Group .......................................................15
2. Regression Analyses for Total Excess Moves ...............................................25
3. Regression Analyses for Total Initial Time ...................................................26
4. Regression Analyses for Total Execution Time ............................................26
5. Regression Analyses for Total Completion Time ..........................................27
6. Welch’s T-Test and Mean Scores by Age Group ........................................29
LIST OF FIGURES

FIGURE

1. Format of Computerized TOL\textsubscript{DX} Problem..................................................17
2. Mean number of excess moves across all 10 trials for each age group ..........22
3. Mean total initial time across all 10 trials for each age group...................22
4. Mean total execution time across all 10 trials for each age group..............23
5. Mean total completion time across all 10 trials for each age group ..........24
CHAPTER I

INTRODUCTION

Starting a fire, building a shed, or replacing a burnt out light bulb all share a need to follow a certain set of steps in order to complete the task. Before beginning any of these tasks it is important to first plan out what steps need to be completed and when. It would not be effective to put up siding before the framework of a shed is complete or to throw logs onto a lit match. Planning can be thought of as the ability to think about and create a mental image of a problem that needs solving (Unterrainer & Owen, 2006). Once this image has been created, various problem solving strategies can be planned out and their consequences analyzed, all before even beginning to actively complete the task. Planning is a central component of executive functioning, which is a set of mental processes that allows for the conscious, goal-directed control of behavior. Given its importance in daily functioning, numerous researchers and psychologists are interested in studying planning and executive functioning. The Tower of London (TOL) is a popular test created for this purpose.

Origins and Development of the Tower of London

The TOL is a commonly used tower transfer task originally designed to detect and measure planning deficits in individuals with frontal lobe lesions. In the manual TOL, participants are shown two identical stimulus boards containing three colored spheres; red, blue, and green. These are mounted on three wooden pegs of varying size able to hold up to one, two, or three spheres. The arrangement in which each trial starts varies
depending on the version of the TOL that is being given. Some versions have a single start state that is displayed at the beginning of each trial and others utilize a novel start state at the beginning of each trial. The participant’s goal is to match their arrangement (start state), to the arrangement displayed on the experimenter’s board (goal state).

Measures of performance on the TOL vary between studies, but the most commonly used metrics of success are trial completion (creating the correct arrangement), total number of moves, total execution time, and total completion time (Berg & Byrd, 2002; Berg, Byrd, McNamara, & Macdonald, 2006).

The TOL was adapted from another tower transfer task, the Tower of Hanoi (TOH) (Shallice, 1982). The TOH requires participants to transfer an entire stack of three discs of decreasing size from one peg to another in as few moves as possible without stacking a larger disc on top of a smaller one. Both of these tests are similar in the sense that the participant must use their planning abilities to find and execute the most efficient means of moving a series of spheres or discs from a starting state to a goal state that has been specified by the examiner (Berg & Byrd, 2002). However, in recent years the TOL has become more commonly used than the TOH due to its wider and more versatile problem set and the greater number of testing parameters that are measured (Rahm, Köstering, & Unterrainer, 2011).

The TOL has a diverse problem set that exposes participants to a wide variety of different tower arrangements. Various forms of the TOH have utilized slightly different arrangements however, as soon as the “trick” to solving the problem is discovered individuals simply need to refine this strategy in subsequent trials. There is no iterative strategy underlying successful solution of TOL problems. Thus, the individual must
utilize a greater assortment of different strategies for each trial. There may be some overlap in the strategies used to solve TOL problems as some problem sets utilize problem nesting, which is the process of embedding moves from a 3 move problem into a 5 move problem later in the problem sequence (Berg & Byrd, 2002). Nesting a problem allows the researchers to add additional moves that will place a greater strain on the individual’s working memory while retaining similar planning issues or strategies. Additionally, different TOL arrangements allow the experimenter greater control over how difficult the problem set will be and to systematically increase problem difficulty. As well as controlling for difficulty, the experimenter can create a wider variety of qualitatively different problems (Berg & Byrd, 2002; Shallice & Burgess, 1991). Having a wider variety of potential problems gives the TOL greater applicability to a variety of populations with varying levels of impairment. Since there are numerous arrangements of spheres and pegs available, experimenters are able to control the difficulty of their test as the problem sequence progresses. This allows experimenters to tailor the test specifically to the population of interest (Kaller et al., 2012; Unterrainer & Owen, 2006). For example, Köstering et al. (2015) found that participants who were in a cognitively impaired sample scored consistently lower than the participants in the healthy sample population. If a researcher were to develop a test for this type of population, they may consider creating a more simplistic problem set so the hardest problem is easier than in a normal problem set.

When Shallice originally created the TOL he proposed that inhibition would be an important contribution to performance. When working on a problem, individuals are likely to use what he referred to as Contention Scheduling (Shallice, 1982). He proposed
that contention scheduling is a fast and automatic process by which an individual will quickly view and process an unfamiliar set of visual stimuli and the mind will rapidly select what it believes to be the most relevant schema to fit the current situation. In relation to the TOL, a tower arrangement will be viewed and the mind will select the schema, or set of actions, that it believes to be most efficient for that problem. The strongest and most readily accessible schema is also the most familiar. The reason Shallice (1982) thought inhibition was so important was because as participants progress through the test they are exposed to more and more problems and begin to formulate numerous solution schemata. Though the trials may be similar in appearance, the optimal solution may be different and the participants must inhibit non-optimal schemata from becoming activated, which could lead to an incorrect or inefficient problem solution. Shallice cited this as one of the main reasons that individuals with frontal lobe lesions performed poorly on the TOL. They were not able to take into account all of the visual stimuli in each problem and as a result inefficient schemata were activated and followed when solving various problems.

**TOL-Drexel 2nd Edition**

There are many different forms of the TOL however, the present study will focus on a computerized version of the Tower of London-Drexel 2nd Edition (TOL\textsubscript{DX}). The manual version of the test was originally adapted from the Shallice TOL by Culbertson and Zillmer (1998) as a test of executive function in children and was later modified to assess executive functioning in adults with attention deficit hyperactivity disorder (ADHD) (Riccio, Wolfe, Romine, Davis, & Sullivan, 2004). The manual TOL\textsubscript{DX} is a commercially available test, consisting of 10 trials ranging in moves from 4 to 7.
Although many of the rules remain the same, several changes were made to improve upon the original TOL. The first of these changes was to remove repeated trials which were used to detect how many times a participant would need to try a problem before they solved it using the minimum number of moves or, as it will be referred to here, the *optimal path*. By eliminating the repeated trials the test provided a higher level of novelty, reduced practice effects, and shortened testing time (Riccio et al., 2004).

The next step in the development of the TOL$_{DX}$ was to include 6 and 7 move problems to reduce the ceiling effect and increase test sensitivity. In their studies with the TOL Kaller, Unterrainer, and Stahl (2012) reported that the use of 7 move problems helps to identify the “upper range” of planning ability. Albert and Steinberg’s (2011) findings that the greatest difference in planning ability was found on the most difficult problems with the greatest number of moves further support this concept. The use of higher move problems allows the users of the TOL$_{DX}$ to differentiate between high and low planning performance.

Arguably, the most significant improvement in the TOL$_{DX}$ was the standardization of the test and the development of a comprehensive normative base. A detailed set of instructions for administration, scoring, and interpretation were created to reduce the variability among administrations of the TOL$_{DX}$ and to increase the comparability of studies (Culbertson, Moberg, Duda, Stern, & Weintraub, 2004; Culbertson & Zillmer, 2001). A common thread in the TOL literature is that a variety of different forms of the TOL have been used (“balls and pockets” and different problem sets) and performance is analyzed using different outcome measures (initial time and excess moves vs total trial completions and rule violations). As the inter-test variability increases, it becomes
progressively more difficult to compare findings since different parameters tap into different cognitive processes. By standardizing the procedure researchers are able to investigate the same cognitive processes with different populations, showing how various conditions can differentially affect planning ability.

After creating the test and the administration guide, Culbertson and Zillmer (2001) established a sizeable lifespan normative base. The TOL\text{DX} was normed using a sample of American and Canadian individuals who demonstrated an average or above level of cognitive and physical functioning. These individuals ranged from 7 to 80 years of age ($n = 990$). This sample was then broken into nine different age brackets; 7-8 ($n = 110$), 9-10 ($n = 157$), 11-12 ($n = 103$), 13-15 ($n = 76$), 16-19 ($n = 162$), 20-29 ($n = 192$), 30-39 ($n = 74$), 40-59 ($n = 77$), and 60-80 ($n = 39$). Norms were also established for children with Attention Deficit Hyperactivity Disorder (ADHD) using two different clinical child populations ranging in age from 7-15 ($n = 129; n = 115$). Culbertson and Zillmer (2001) reported norms on a set of outcome measures used in the TOL\text{DX}, further increasing comparability across studies. These measures include total move score (total number of moves to solve a problem), total correct score (number of problems solved using optimal path), total time violations (number of times participant exceeds 1 and 2 minute time limits), total rule violations (total move errors committed on a problem), total initiation time (time between first exposure to problem and first move), total execution time (time spent physically moving spheres after the initial think time), and total problem solving time (total completion time across 10 trials). The present study will not address errors and will focus solely on the total move score and the various completion times.
CHAPTER II

REVIEW OF THE LITERATURE

Although the TOL\textsubscript{DX} was developed as a manual test to be implemented using wooden stimulus boards, there is an increasing amount of evidence to support the use of computerized methods of testing. Using computerized testing methods allows for greater standardization of testing procedures and can reduce experimenter biases, such as an experimenter unintentionally flinching or grimacing at an incorrect move. Additionally, since the test has been standardized, experimenters do not need to be as highly trained as would be necessary in a traditional testing format (Zygouris & Tsolaki, 2015). Research assistants only need minimal amounts of training. As well as ease of administration, computerized testing simplifies the data gathering process and in many cases can increase the accuracy of the measurement. Since the administration and data gathering and analysis all take place on the same platform, the data can be stored between tests and changes in performance can be easily tracked over time (Cambridge Cognition, 2012; CNS Vital Signs, 2012; Neurotrax Corporation, 2003;).

There are numerous benefits to using computerized testing, but there are also risks that come along with the use of technology. One of the biggest concerns is the utility and validity of computerized tests in the screening of older adults. Older adult populations may not be as familiar with some of the interfaces used in these studies (Zygouris & Tsolaki, 2015) and their computer illiteracy can cause anxiety and create confounds in the data (Tierney & Lermer, 2012). Though these are all valid risks that must be taken into
account when designing a study, there is evidence that this population is becoming more computer literate. Hart, Chapparo, and Halcomb (2008) report that older adults are the fastest growing demographic of internet users and younger adults are aging in an increasingly technological world. The older adults of the future will be increasingly familiar with the type of technology used in testing and taking the steps now to establish psychometrically sound measures is increasingly important.

When creating a computerized test it is not always a process of simply taking the exact same test and putting it into a digital format. When a test is converted, there is a chance that the outcome measures of the manual and computer forms of a test may represent different constructs and tap into different abilities. When the two versions do not have high construct validity the original normative base cannot be used and new norms must be established and its psychometric properties reassessed (Gates & Kochan, 2015).

**TOL Background**

Since its creation as a means of assessing planning ability, the general TOL test has been implemented as a tool for measuring various domains of executive functioning (Bottari et al., 2009; Köstering et al., 2015; Owen, 2005) as well as spatial planning (Berg & Byrd, 2005; Berg, Byrd, McNamara, & MacDonald, 2006; Kaller, Unterrainer, & Stahl, 2012; Pulos & Denzine, 2005; Shallice, 1982; Unterrainer & Owen, 2006), working memory (Albert & Steinberg, 2011; Berg & Byrd, 2002; Pulos & Denzine, 2005), inhibition (Albert & Steinberg, 2011; Berg & Byrd, 2002; Shallice, 1982), and task shifting (Pulos & Denzine, 2005).
In addition to Shallice’s (1982) original findings of deficits in individuals with frontal lobe lesions, there have emerged several additional populations that frequently show deficits in their performance on the TOL; those with neurological and psychiatric conditions such as depression or Parkinson’s (Jacobs & Anderson, 2002), individuals who have suffered a traumatic brain injury (TBI), Alzheimer’s and related dementias (Carlin et al., 2000), ADHD (Culbertson & Zillmer, 1998), Autism (Wisley & Howlin, 2009), schizophrenia (Landua & Morris, 2011), and older adults (Albert & Steinberg, 2011; Berg, Byrd, McNamara, & Case, 2010; Köstering, Stahl, Leonhart, Weiller, & Kaller, 2014; MacLeod & Kliegel, 2005).

In previous studies, the TOL$_{DX}$ was shown to differentiate between healthy and clinical populations. Culbertson et al. (2004) conducted a study using 65 patients in an outpatient setting who had been seeking assistance for Parkinson’s disease (PD) and 34 healthy, demographically matched controls. They found the TOL$_{DX}$ was an effective means of detecting executive function deficits in individuals with PD who required a greater numbers of moves and more time to solve each problem than controls. This finding remained even after they controlled for motor deficits. A similar finding was reported by Krishnan, Smith, and Donders (2012) in their study of TOL$_{DX}$ performance with adults who suffered a TBI. Individuals with a TBI, regardless of severity, scored reliably worse than the healthy control group. Deficits were mostly expressed through an increase in the number of moves and execution time, as opposed to overall time, and were posited to be due to the use of inefficient problem solving strategies.

As the literature on the TOL has grown, researchers have discovered several factors that have a significant effect on performance. These factors include age,
arrangement of the boards during a test, and execution strategies used to solve the problems.

Age has repeatedly been shown to be a moderator of performance on all measures of planning ability on the TOL (Albert & Steinberg, 2011; Berg, Byrd, McNamara, & Case, 2010; Köstering, Stahl, Leonhart, Weiller, & Kaller, 2014; MacLeod & Kliegel, 2005). Various studies have found that on average, older adults (age > 60) had a greater number of *total moves* and *excess moves* per trial and they spent a greater amount of time planning than younger participants (Andres & van der Linden, 2000; Bugg, Zook, DeLosh, Davalos, & Davis, 2006). Additionally, Köstering et al. (2014) found an age-related decline in planning performance, as demonstrated by a decrease in the total number of problems solved. They found that this age effect was significant even when the sample size was decreased and compared to other older adults within a 29 year age range (ages 60-89).

**Planning Ability**

There are several cognitive processes that are considered important to planning and goal completion. These processes include 1) the ability to recognize what the goal state looks like and acknowledging when it has been achieved, 2) anticipating the outcome of various behaviors and actions in relation to goal-attainment (Carlin et al., 2000; Kaller et al., 2004), 3) generating and storing mental representations of the process of events that leads from the initial action to goal completion (Unterrainer & Owen, 2006), and 4) the inhibition of distracting and irrelevant stimuli and preventing impulsive, unplanned responses (Shallice, 1982; Unterrainer & Owen, 2006). Planning requires one to create mental images and representations of the problem, the moves necessary to
complete the goal, and being able to mentally extrapolate how these moves will affect the overall goal process. As a result of this, one of the most important aspects of planning is a highly functioning working memory.

Working memory (WM) is responsible for creating the mental representation, or problem space, where the various components for TOL goal completion are generated and, more importantly, organized into a cohesive order (Pulos & Denzine, 2005). Creating a problem space assists in planning the optimal path and visualizing the effect that incorrect moves can have on the problem solution (Berg & Byrd, 2002). In keeping with the importance of inhibition, Baddeley (1996) stated that a common aspect of WM is the capacity to focus attention on a specific task, despite the presence of irrelevant stimuli and interference. Performance on the TOL loads heavily on WM and participants must utilize their WM to create a problem solution and assess how each move could positively or negatively affect goal attainment.

Several studies of WM have shown that it develops rapidly during adolescence, plateaus during adulthood, and eventually decreases in older age (Hedden & Gabrieli, 2004; Raz & Kennedy, 2009). Numerous studies have supported this pattern of WM development and deterioration with the greatest changes in performance happening in early to mid-adolescence and the most significant deterioration beginning in old age (ages 50-60) (Albert & Steinberg, 2011; Alloway & Alloway, 2013; Cowan, Naveh-Benjamin, Kilb, & Saults, 2006; Fandakova, Sander, Werkle-Bergner, & Shing, 2014).

In order to effectively create mental representations, one must first accurately observe and encode the environmental and problem-based stimuli using visuospatial processing abilities. Visuospatial processing is the ability to observe objects within an
environment and assess their spatial relationship to one another (Joyce & Robbins, 1991; Morice & Delahunty, 1996). Visuospatial processing has been closely tied with TOL performance and general planning ability with completion time being highly correlated with visuospatial WM (Gilhooly et al., 2002; Pulos & Denzine, 2005). Gilhooly, Wynn, Phillips, and Della Sala (2002) showed that TOL performance had a high factor loading on visuospatial WM. Individuals with poor visuospatial processing and WM ineffectively differentiate the novel aspects of the current problem from previously viewed trials (Pulos & Denzine, 2005). Similar, yet novel, trials could not be distinguished from previously completed trials and previous schemata and non-optimal problem solutions will be activated.

The processes of visual processing, spatial orientation, and speed of processing have been grouped into one overarching construct called fluid cognitive skills (also considered fluid intelligence), and all are susceptible to a decline during normal, healthy aging (Finkel, Reinold, McArdle, & Pederson, 2003; McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002). As discussed, WM and visuospatial processing are important to the TOL_{DX} solution since they are essential components to the creation and maintenance of the mental representations of the solutions. The present study investigated the effect that visuospatial ability, fluid intelligence, and WM would have on the number of excess moves, initial time, execution time, and total completion time across all ten trials on a computerized version of the well-established manual TOL_{DX} test.

Construct and criterion-related validity has been established for the manual TOL_{DX} through comparisons with other problem solving, neuropsychological, and cognitive tests (Culbertson and Zillmer, 1998a; 2001; Larochette, Benn, & Harrison,
Riccio et al. (2004) found significant correlations between TOL$_{DX}$ performance (decreased moves and completion times) and processing speed, perceptual skills, matrix reasoning, and immediate memory. The TOL$_{DX}$ was found to have a weak negative correlation with scores of general intelligence (Culbertson & Zillmer, 2001; Culbertson et al., 2004; Sari & Culbertson, 2001), which would indicate that the TOL$_{DX}$ is a measure of the specific executive function of planning, as opposed to a measure of overall intelligence. While greater intelligence does not likely mean poor TOL performance through greater numbers of moves or completion times, the TOL is simply not an accurate measure of overall intelligence as it more specifically measures planning ability.

The current lifespan study sought to further demonstrate the age effects that are present in measures of performance on the TOL and the constructs that are being assessed. Since the creation of the TOL$_{DX}$, only one additional study has created a semi-lifespan normative base however, the participants ($n = 344$) ranged in age from 50-90 years of age (Pena-Casanova et al., 2009). It is hypothesized that on a computerized version of the TOL$_{DX}$, age will have an effect on excess moves, initial time, execution time, and total completion time. The present study will add to the literature and provide a base of normative data on a computerized version of the TOL$_{DX}$.

The *Frontal Hypothesis of Cognitive Aging* suggests that aging primarily affects behaviors that are mediated by the frontal lobes such as general executive functioning (Moscovitch & Winocur, 1995; West, 1996). Given that planning is an important component of executive functioning, the present study sought to further support this theory as it pertains to a computerized version of the TOL$_{DX}$ (Phillips, MacLeod, & Kliegel, 2005). It was expected that age will have an effect on TOL completion with
performance being lowest (greatest excess moves and completion times) in adolescents and older adults and peak performance occurring in early adulthood.

The following outcomes were hypothesized:

H₁: Age will have a negative effect on all outcome measures with older adults and adolescents having the greatest number of excess moves, execution time, and completion times as well as the lowest initial times.

H₂: In a multiple regression, all factors (age, sex, years of education, WM, fluid intelligence, and visuospatial performance) will have a positive effect on the outcome measures; total excess moves, total initial time, total execution time, and total completion time.
   • With the exception of age, the greater the presence of the construct, the better the TOL performance

H₃: Of the 5 factors:
   • Sex and years of education will have the smallest effect
   • Age will have the greatest negative effect
   • WM, fluid intelligence, and visuospatial processing will have a smaller effect than age and more of an effect than sex and years of education.

H₄: There is minimal research available on the adaptation of a computerized version of the TOLDX and the present study performed exploratory analyses to investigate the relationship between the manual and computerized TOLDX.
CHAPTER III

METHOD

Participants

The majority of participants were drawn from an active database currently consisting of over 7,200 individuals. The participants \((n = 930)\) ranged from 16 to 80 years of age. The lower and upper bounds of this age range were based on the cutoffs used by Culbertson and Zillmer (2001) in their creation of the adult version of the TOL\(_{DX}\). The participants were divided into 5 different conditions based on the age groups used in the original manual TOL\(_{DX}\) study to increase comparability of the two tests.

Participant age groups were the following: 16-19 \((n = 319)\), 20s \((n = 390)\), 30s \((n = 78)\), 40-59 \((n = 65)\), and 60-80 \((n = 78)\). The mean age, years of education, and the proportion of males and females for each age group are shown in Table 1.

Table 1

Demographics for Each Age Group

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Sample Size</th>
<th>M ± SD</th>
<th>Mean Years Edu.</th>
<th>% Male/Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>16-19</td>
<td>319</td>
<td>18.44 ± 1.31</td>
<td>12.95 ± 1.05</td>
<td>27.7/72.3</td>
</tr>
<tr>
<td>20s</td>
<td>390</td>
<td>22.65 ± 2.79</td>
<td>14.72 ± 1.18</td>
<td>35.8/64.2</td>
</tr>
<tr>
<td>30s</td>
<td>78</td>
<td>33.59 ± 3.29</td>
<td>15.07 ± 1.28</td>
<td>39.7/60.3</td>
</tr>
<tr>
<td>40-59</td>
<td>65</td>
<td>48.17 ± 5.56</td>
<td>15.15 ± 1.41</td>
<td>36.9/63.1</td>
</tr>
<tr>
<td>60-80</td>
<td>78</td>
<td>69.01 ± 5.41</td>
<td>15.13 ± 2.00</td>
<td>33.3/66.7</td>
</tr>
<tr>
<td>Total</td>
<td>930</td>
<td>27.85 ± 15.10</td>
<td>14.21 ± 1.56</td>
<td>33.2/66.8</td>
</tr>
</tbody>
</table>

Participants were recruited from the University of Colorado at Colorado Springs psychology classes; older adults were recruited from the Gerontology Center database; 16
and 17 year old participants, as well as additional older adult participants, were recruited from relatives of the student and older adult participants.

Participants were excluded from participation in the study if they reported any current or past neurological disease (schizophrenia, dementia, etc.), traumatic brain injury, learning disability, major psychiatric condition (depression, generalized anxiety disorder, etc.) or were currently using a substance they believed to have an effect on their cognition; either illicit or prescribed (Ritalin, cannabis, etc.).

Sixteen and 17 year old participants received $5 per hour for their participation. College students received extra credit or SONA research credits for each hour of participation. SONA credits are awarded to students for participation in campus based research. Many undergraduate psychology courses require a minimum number SONA credits as a component of the students overall grade. Older adults (age 60 and older) received $10 for each hour of participation. Each testing session was limited to a maximum of two hours per day. Testing sessions were capped at two hours to reduce any variance that might arise from participant fatigue.

**Instruments**

Prior to participation in this study, all participants completed an extensive demographic and general health questionnaire and signed an informed consent.

Participants were administered a computerized version of the TOL\textsubscript{DX} in a quiet testing room. The computerized TOL\textsubscript{DX} was presented on a desktop computer and was manipulated using either a mouse or touch screen. A computerized schematic of a typical problem is shown in Figure 1.
Participants were administered a computerized 10 trial TOL_DX problem set that ranged from 4 to 7 required moves and were allowed a maximum of 20 moves or 2 minutes in which to complete the trial. Prior to beginning the test, the rules were presented with an initial practice problem, followed by two additional practice trials. The rules were visually presented and read in a digitized voice as follows:

See these two boards? They are both alike. This board on the left is the one you will use and you will make it look like the one on the right. Your task is to make this arrangement on the left look like the one of the right in as few moves as possible. There are two rules you must follow when you are arranging the beads. The first rule is that you are not allowed to place more beads on a peg than it will hold. The second rule is that you can only move one bead at a time. You cannot move two beads off the pegs at the same time. Do you have any questions? Now, arrange the beads on the left so they look like the arrangement on the right. You have two minutes to do each problem. Also, you must complete the problem in 20 moves or fewer. If you do not finish in two minutes or in less than 20 moves, the trial ends and a new problem is presented.

Participants were given the opportunity to take part in a larger test battery. Tests from this battery that were included in the analyses were the n-Back, visual span, and the Wechsler Abbreviated Scale of Intelligence (WASI) matrix subtest.

The alphabetical n-Back test is an assessment of WM. The n-Back is a computerized test in which the participant is shown a series of letters and they must
report whether or not they have seen that same letter, one, two, or three letters back (i.e. A, B, F, A, B, G, etc.). To begin the test, the participant is shown 10 practice letters. Once the practice has been completed, the participants are shown 100 letters for two seconds at a time, with two seconds between each letter. Of the 100 letters, there are 30 targets that the participants attempt to identify as targets and 70 distractors that must be identified as non-targets. Prior to beginning the 3-back that was used for the analyses, participants completed 100 trials of both the 1- and 2-back. The total number of correct responses on the 3-back trial was used to assess WM.

Participants’ visuospatial ability was assessed using a computerized visual span-backwards test. During this test, participants are seated in front of a computer and eight blue boxes are randomly arrayed across the screen. One at a time, in a sequential order, the boxes flashed and turned white for one second at a time. The test begins with sequences in which two blocks flash and finishes with eight-block sequences. Each sequence was presented twice and the test was discontinued if the participant failed to reproduce two sequences of the same length. The tests contains two parts; the visual span forward followed by the visual span backward. In the visual span-backwards test, the participant must correctly select, using a mouse, each of the boxes that were illuminated in the reverse order in which they flashed (i.e. the first box to flash will be the last that is clicked by the participant and vice versa). The total number of backward correct responses was used in the analyses.

Fluid intelligence was represented by the raw performance score on the Matrix subtest of the Wechsler Abbreviated Scale of Intelligence-1st Edition (WASI). During this test, the participant is seated across from an examiner and presented with images with
increasingly difficult patterns. In each image or series of images, there is one portion of the image missing and the participant must select from presented options the image which most accurately completes the picture. The test was discontinued after four incorrect responses in a row. Raw scores were used on the Matrix test to increase comparability of the scores across the age groups since they were not age-adjusted prior to comparison.

Following the initial information gathering, all participants completed a variety of cognitive and neuropsychological tests. Each one to two hour testing session included a variety of tests that assessed memory, problem solving, inhibition, attention, task switching ability, and general intelligence. These additional tests were presented in a random order and not all participants completed the same tests. There was a minimum of two days between testing to reduce potential practice effects.

When completing the TOL\textsubscript{DX} the participant was brought into a quiet room and seated in front of a computer. The previously stated rules were presented on the computer prior to beginning the test and any questions were answered. The TOL\textsubscript{DX} is a combination of a time and performance limited test in which participants were allowed a maximum of 20 moves or 2 minutes per trial. If they failed to complete the trial within the set parameters, the time and total number of moves were recorded as the highest possible value and they move to the next trial, as per the instructions provided by Culbertson and Zillmer (2001) in their TOL\textsubscript{DX} manual. When a participant exceeded the 2 minute or 20 move limit, the number of participant was given the maximum number of excess moves which was the minimum number of moves possible for that problem subtracted from 20; i.e. a time violation on a 4 move problem will be recorded as 16
excess moves. The participants were presented with two practice problems to learn the rules of the TOL\textsubscript{DX}, followed by the ten trials that make up the test itself.

All timing and recording of moves were recorded digitally by the testing software and manually entered into the data set. The initial/think time, execution time, total time, and the total number of excess moves were recorded and analyzed in SPSS Statistics Version 24. Each of these measures were totaled across all ten trials of the TOL\textsubscript{DX}. The independent variables that were assessed across these trials was age, sex, years of education, fluid intelligence, WM, and visuospatial ability. All t-tests were run using QuickCalcs from GraphPad Software due to the absence of Welch’s t-test on SPSS Statistics version 24 (available from https://www.graphpad.com/quickcalcs/ttest1.cfm).

Since not all participants completed the same number of testing sessions or stayed for the same amount of time, not everyone who completed a computerized TOL\textsubscript{DX} also completed an n-Back, Matrix, or visual span test. However, a select cases process was used in SPSS to ensure that of those who completed an n-Back, Matrix, and TOL\textsubscript{DX}, only those who had also completed a computerized TOL\textsubscript{DX} were included in the analyses.
CHAPTER IV

RESULTS

The present study investigated the effect of age, sex, years of education, fluid intelligence as assessed by performance on the Matrix subtest of the Wechsler Abbreviated Scale of Intelligence-1st edition (WASI), working memory, and visuospatial ability, on total excess moves, total initial time, total execution time, and total time. The comparability of a computerized and manual version of the TOL\textsubscript{DX} were also investigated. Time and excess move values were totaled across all 10 trials of the TOL\textsubscript{DX}. All analyses were run through SPSS Statistics version 24.

ANOVA and Post Hoc

An Analysis of Variance (ANOVA) was conducted to detect the effect of age on performance. Age was shown to have a significant effect on total excess moves such that, as age increased, participants needed a greater number of moves to solve each trial, $F(4, 643) = 7.18$, $p < .001$. The mean number of total excess moves for each group are displayed in Figure 2.

Post hoc comparisons using the Sidak test indicated that there were significant differences between the numbers of moves required to complete the problem for each age group. 16-19 year olds needed significantly more moves ($M = 27, SD = 16.63$) than the 20 year olds ($M = 20.65, SD = 15.14$) and 60-80 ($M = 31.24, SD = 19.82$) year olds needed significantly more moves than both the 30 year olds ($M = 21.33, SD = 18.01$) and 20 year olds ($M = 20.65, SD = 15.14$).
Figure 2. Mean number of total excess moves across all 10 trials for each age group.

Age was also shown to have a significant effect on the total initial time such that, as age increased, participants spent a longer amount of time planning and thinking prior to their first move, $F(4, 645) = 7.21, p < .001$. The mean total initial time for each age condition are shown in Figure 3.

Figure 3. Mean total initial time in seconds across all 10 trials for each age group.

Post hoc comparisons using the Sidak test indicated that 16-19 year olds ($M = 95.66, SD = 56.02$) spent a significantly shorter time planning than all of the other age groups; 20s ($M = 111.06, SD = 60.19$), 30s ($M = 120.52, SD = 54.51$), 40-59 ($M = 130.67, SD = 68.82$), and 60-80 ($M = 132.80, SD = 61.65$). No significant differences existed between any of the other groups.
Age was also shown to have a significant effect on the total execution time such that, as age increased, participants spent a longer amount of time executing the solution and physically completing the problem, $F(4, 633) = 43.19, p < .001$. The mean total execution times for each age condition are shown in Figure 4.

![Figure 4](image.png)

*Figure 4. Mean total execution time across all 10 trials.*

Post hoc comparisons using the Sidak test revealed that older participants required greater amounts of time to execute the problem than younger participants. 60-80 year olds ($M = 360.65, SD = 150.30$) took significantly longer than all other age groups; 40-59 ($M = 255.45, SD = 102.46$), 30s ($M = 213.04, SD = 89.18$), 20s ($M = 187.35, SD = 77.88$), and 16-19 ($M = 206.24, SD = 79.79$). Additionally, 40-59 year olds took significantly longer than all younger groups except for 30 year olds. Significant differences did not exist between the younger age groups.

Age was shown to have a significant effect on total completion time such that, as age increased, participants needed a greater amount of time to complete the trial, $F(4, 634) = 41.6, p < .001$. The mean total completion times for each group are displayed in Figure 5.
Figure 5. Mean total completion time in seconds across all 10 trials for each age group.

Post hoc comparisons using the Sidak test showed that 60-80 year old participants ($M = 493.45$, $SD = 147.07$) took significantly longer to complete the test than all other age groups; 40-59 ($M = 385.42$, $SD = 108.44$), 30s ($M = 333.26$, $SD = 99.41$), 20s ($M = 299.12$, $SD = 106.37$), and 16-19 ($M = 302.02$, $SD = 98.98$). Similar to the execution time, 40-59 year olds took significantly longer than all younger groups except for the 30 year olds. There were no significant differences in any of the younger age groups.

Regression Analyses

A bivariate regression analysis was conducted to investigate the relationship between the total number of excess moves and the six predictors; visual span backward, years of education, n-Back (3-back) correct responses, sex, age, and the raw score on the Matrix sub test. There was a significant relationship between the number of excess moves and the six predictors and the overall model accounted for 17% of the variance on total excess moves, $F(6, 85) = 2.89$, $p = .01$, $R = .41$, $R^2 = .17$, $R^2_{\text{adj}} = .11$. There was only one significant standardized $\beta$ coefficient in the model. Total excess moves was uniquely predicted by fluid intelligence, as measured by performance on the Matrix subtest of
WASI-I, $b = -1.25, \beta = -0.26, p = 0.03$. Complete results for the multiple regression are presented in Table 2.

Table 2

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Unstandardized B</th>
<th>Standard Error B</th>
<th>Standardized Beta (β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.07</td>
<td>0.11</td>
<td>0.07</td>
</tr>
<tr>
<td>Sex</td>
<td>3.11</td>
<td>4.5</td>
<td>0.07</td>
</tr>
<tr>
<td>Years of Education</td>
<td>1.14</td>
<td>1.30</td>
<td>0.09</td>
</tr>
<tr>
<td>Fluid Intelligence</td>
<td>-1.25*</td>
<td>0.58</td>
<td>-0.26*</td>
</tr>
<tr>
<td>Working Memory</td>
<td>0.05</td>
<td>0.29</td>
<td>0.02</td>
</tr>
<tr>
<td>Visuospatial Ability</td>
<td>-1.62</td>
<td>1.35</td>
<td>-0.15</td>
</tr>
</tbody>
</table>

* $p < .05$

An additional bivariate regression analysis was conducted to examine the effect of the six predictors on the total initial time on the TOL$_{DX}$. There was a significant relationship between the total initial time and the six predictors and accounted for 17% of the variance on total initial time, $F(6, 85) = 2.85, p = .01, R = .41, R^2 = .17, R^2_{ADJ} = .11$. There were only two significant standardized $\beta$ coefficients in the model. Total initial time was uniquely predicted by increased age, $b = 1.16, \beta = .40, p = .001$, and visuospatial ability as measured by performance on the visual span backward test, $b = 10.15, \beta = .30, p = .01$. Complete results for the multiple regression are presented in Table 3.

A third bivariate regression analysis was conducted to examine the effect of the six predictors on the total execution time on the TOL$_{DX}$. There was a significant relationship between the total execution time and the six predictors and accounted for 64% of the variance on total initial time, $F(6, 64) = 19.16, p < .001, R = .80, R^2 = .64, R^2_{ADJ} = .61$. There were only two significant standardized $\beta$ coefficients in the model. Total initial time was uniquely predicted by increased age, $b = 3.28, \beta = .44, p < .001$, and
and fluid intelligence, \( b = -11.32, \beta = -.32, p = .001 \). Complete results for the multiple regression are presented in Table 4.

Table 3

*Regression Analyses for Total Initial Time*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Unstandardized B</th>
<th>Standard Error B</th>
<th>Standardized Beta (( \beta ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1.16**</td>
<td>.34</td>
<td>.40**</td>
</tr>
<tr>
<td>Sex</td>
<td>-9.09</td>
<td>13.57</td>
<td>-.07</td>
</tr>
<tr>
<td>Years of Education</td>
<td>-5.10</td>
<td>3.91</td>
<td>-.14</td>
</tr>
<tr>
<td>Fluid Intelligence</td>
<td>1.70</td>
<td>1.76</td>
<td>.12</td>
</tr>
<tr>
<td>Working Memory</td>
<td>.29</td>
<td>.86</td>
<td>.04</td>
</tr>
<tr>
<td>Visuospatial Ability</td>
<td>10.15**</td>
<td>4.07</td>
<td>.30**</td>
</tr>
</tbody>
</table>

** \( p < .01 \)

Table 4

*Regression Analyses for Total Execution Time*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Unstandardized B</th>
<th>Standard Error B</th>
<th>Standardized Beta (( \beta ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>3.28**</td>
<td>.65</td>
<td>.44**</td>
</tr>
<tr>
<td>Sex</td>
<td>-11.89</td>
<td>26.91</td>
<td>-.03</td>
</tr>
<tr>
<td>Years of Education</td>
<td>7.59</td>
<td>7.76</td>
<td>.08</td>
</tr>
<tr>
<td>Fluid Intelligence</td>
<td>-11.32**</td>
<td>3.34</td>
<td>-.32**</td>
</tr>
<tr>
<td>Working Memory</td>
<td>-3.70</td>
<td>1.90</td>
<td>-.16</td>
</tr>
<tr>
<td>Visuospatial Ability</td>
<td>-12.17</td>
<td>7.55</td>
<td>-.15</td>
</tr>
</tbody>
</table>

** \( p < .01 \)

The final bivariate regression was conducted to examine the effect of the five predictors on the total completion time. There was a significant relationship between the total completion time and the six predictors and accounted for 56% of the variance on total completion time, \( F(6, 84) = 17.81, p < .001, R = .75, R^2 = .56, R^2_{ADJ} = .53. \) There were only two significant \( \beta \) coefficients in the regression model. Total completion time was uniquely predicted by increased age, \( b = 4.27, \beta = .57, p < .001 \), and fluid intelligence, \( b = -10.5, \beta = -.28, p = .003 \). Complete results for the multiple regression are presented in Table 5.
Table 5

_Regression Analyses for Total Completion Time_

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Unstandardized B</th>
<th>Standard Error B</th>
<th>Standardized Beta (β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>4.27***</td>
<td>.66</td>
<td>.57***</td>
</tr>
<tr>
<td>Sex</td>
<td>-5.64</td>
<td>26.23</td>
<td>-.02</td>
</tr>
<tr>
<td>Years of Education</td>
<td>-2.54</td>
<td>7.53</td>
<td>-.03</td>
</tr>
<tr>
<td>Fluid Intelligence</td>
<td>-10.50**</td>
<td>3.39</td>
<td>-.28**</td>
</tr>
<tr>
<td>Working Memory</td>
<td>.25</td>
<td>1.65</td>
<td>.01</td>
</tr>
<tr>
<td>Visuospatial Ability</td>
<td>-3.03</td>
<td>7.84</td>
<td>-.03</td>
</tr>
</tbody>
</table>

**p < .01, ***p < .001.

A series of Pearson-r correlations were conducted to further explore the relationship between the 5 predictors and the outcome variables to investigate whether there was a possibility for any effects to be lost due to shared variance. These regressions are listed in Appendix A.

**Welch’s T-Tests**

A series of Welch’s t-tests were run to compare the results of the computerized TOL_Dx with the original norms for the manual TOL_Dx provided by Culbertson and Zillmer (2001). A Welch’s t-test was used since the variance was different between the two different datasets and only the means and standard deviations were provided by Culbertson and Zillmer. A total of five different t-tests were conducted, one for each age condition. Each test compared the four outcome measures previously discussed.

When comparing participants ages 16-19 years of age, the scores between the computerized and manual versions differed significantly across all measures of performance; total move score, $t(391) = 33.54, p < .001$; initial time, $t(391) = 10.89, p < .001$; execution time, $t(387) = 3.96, p < .001$; and total completion time, $t(387) = 8.6, p < .001$. 
When comparing participants in their 20s, the scores between the computerized and manual versions differed significantly on total move score, $t(452) = 31.87, p < .001$; initial time, $t(450) = 9.10, p < .001$; and total completion time, $t(451) = 6.01, p < .001$. The total execution time did not differ significantly between the two versions of the test.

When comparing participants in their 30s, the scores between the computerized and manual versions differed significantly on total move score, $t(129) = 16.10, p < .001$; initial time, $t(129) = 7.91, p < .001$; and total completion time, $t(126) = 4.42, p < .001$. The total execution time did not differ significantly between the two versions of the test.

When comparing participants 40-59 years of age, the scores between the computerized and manual versions differed significantly on total move score, $t(122) = 13.72, p < .001$; initial time, $t(122) = 8.18, p < .001$; and total completion time, $t(121) = 5.19, p < .001$. The total execution time did not differ significantly between the two versions of the test.

When comparing participants 60-80 years of age, the scores between the computerized and manual versions differed significantly on total move score, $t(88) = 12.31, p < .001$; initial time, $t(88) = 13.39, p < .001$; total execution time $t(88) = 2.57$; and total completion time, $t(88) = 4.19, p < .001$.

Results of all t-tests can be seen in Table 6.
Table 6

*Welch’s T-Test Analyses by Age Group*

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Measure</th>
<th>Manual TOL-DX</th>
<th>Computer TOL-DX</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>16-19</td>
<td>Initial Time</td>
<td>41.7</td>
<td>95.8</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>Execution Time</td>
<td>173.0</td>
<td>206.2</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>Total Time</td>
<td>214.7</td>
<td>302.0</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>Move Score</td>
<td>29.0</td>
<td>81.8</td>
<td>.0001</td>
</tr>
<tr>
<td>20s</td>
<td>Initial Time</td>
<td>59.8</td>
<td>111.0</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>Execution Time</td>
<td>175.8</td>
<td>187.3</td>
<td>.1323</td>
</tr>
<tr>
<td></td>
<td>Total Time</td>
<td>235.7</td>
<td>298.32</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>Move Score</td>
<td>27.9</td>
<td>75.7</td>
<td>.0001</td>
</tr>
<tr>
<td>30s</td>
<td>Initial Time</td>
<td>52.8</td>
<td>120.2</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>Execution Time</td>
<td>195.2</td>
<td>187.4</td>
<td>.2745</td>
</tr>
<tr>
<td></td>
<td>Total Time</td>
<td>52.8</td>
<td>333.3</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>Move Score</td>
<td>27.3</td>
<td>77.1</td>
<td>.0001</td>
</tr>
<tr>
<td>40-59</td>
<td>Initial Time</td>
<td>48.8</td>
<td>130.0</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>Execution Time</td>
<td>217.8</td>
<td>255.5</td>
<td>.0723</td>
</tr>
<tr>
<td></td>
<td>Total Time</td>
<td>266.7</td>
<td>385.4</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>Move Score</td>
<td>31.1</td>
<td>81.0</td>
<td>.0001</td>
</tr>
<tr>
<td>60-80</td>
<td>Initial Time</td>
<td>72.0</td>
<td>132.8</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>Execution Time</td>
<td>285.3</td>
<td>360.7</td>
<td>.0119</td>
</tr>
<tr>
<td></td>
<td>Total Time</td>
<td>357.4</td>
<td>493.4</td>
<td>.0001</td>
</tr>
<tr>
<td></td>
<td>Move Score</td>
<td>38.8</td>
<td>86.2</td>
<td>.0001</td>
</tr>
</tbody>
</table>
CHAPTER V
DISCUSSION

The purpose of the present study was to assess the utility of a computerized version of the Tower of London-Drexel 2nd Edition as compared to a manual version using wooden stimulus boards and beads. The study assessed the computerized TOLDX by running a series of analyses to test the criterion validity with results found for the manual TOL. Focus was also placed on the creation of a lifespan normative base for the computerized version, using participants ranging from 16-80 years of age. Lastly, exploratory analyses were conducted to compare the total move score, initial time, execution time and total completion time, on the computerized and manual versions of the test.

It was hypothesized that age would have an effect on all of the outcome measures; total excess moves, total initial time, total execution time, and total completion time. Older adults and adolescents would require the most moves and longest time to complete the problems. Additionally, adolescents and older adults would spend the shortest amount of time planning. It was also hypothesized that in a series of multiple regressions, age, sex, years of education, WM, visuospatial performance, and fluid intelligence would all have an effect on the outcome measures. Of these predictors, it was hypothesized that age would have the greatest effect, followed by WM, visuospatial processing, and fluid intelligence, and sex and years of education would have the smallest effect.
Discussion of Results

The hypothesis that age would have an effect on the outcome measures was supported, with age having a significant effect on all of the outcome measures. In line with past research, older participants required a greater number of moves (Andres & van der Linden, 2000; Bugg, Zook, DeLosh, Davalos, & Davis, 2006) and needed more time to complete the trials (Albert & Steinberg, 2011; Berg, Byrd, McNamara, & Case, 2010). Based on the theories of inhibition, it was predicted that as participants’ age increased, they would be less capable of efficiently inhibiting their automatic responses and would rapidly and impulsively select the first solution that became mentally available to them (Albert & Steinberg, 2011; Culbertson & Zillmer, 2001; Steinberg et al., 2009). However, this hypothesis was not supported and results of the present study (see Figure 3) indicated that the initial time increased until age 80 with only a slight plateau occurring around age 60. This is particularly interesting since older adults had the greatest number of excess moves which, in contrast to past research, would indicate that these longer initial times were not spent planning effective solutions to the problem. More research into what is happening during the prolonged think times and why it is not translating to effective solutions. One way in which this could be done would be through the use of think out loud procedure. This would involve participants verbally narrating their thought process as they complete each phase of the TOL; i.e. initial and execution phases.

The regression analyses did not reveal all of the hypothesized effects with age only contributing variance to initial time, execution time, and total time, but not total excess moves. In support of past research, fluid intelligence contributed variance to nearly all of the outcome variables with the exception of initial time. Visuospatial ability,
which was hypothesized to contribute some of the largest variance, only had an effect on the total initial time and no other outcome variables. Given that all three regression models were significant when taken as a whole, contributing between 17-64% of the variance to their respective outcomes, it was surprising to see that only a few predictors contributed a meaningful amount of variance to the regression models. Given that there were 6 predictors in each regression model, it is possible that some of the effect sizes were reduced due to shared variance.

To test the hypothesis of reduced effect sizes due to shared variance, a series of correlation analyses were conducted. These correlations found a few more significant relationships between the predictors and outcome variables with visuospatial processing and WM having moderately strong relationships to the outcome measures. This was especially the case in regards to the number of excess moves. In the regression analysis, there was only one significant predictor and many of the other values were in an unexpected direction. This may lend support for utilizing the correlational analyses as a more accurate representation of the relationship between the predictors and outcome variables. However, when taken as a whole the correlation analyses did not lead to a greater understanding of the small effect sizes or the potential for reduced effect sizes due to shared variance in the regression models. For correlation values see Appendix 1.

When taken together, the multiple regressions and the correlations provide minimal evidence for the criterion validity of the TOL\textsubscript{DX}. There was some promising data showing a relationship between visuospatial processing, fluid intelligence, and age on computerized TOL\textsubscript{DX} performance. Interestingly, of all the correlations, WM and initial time had the weakest relationship. This was surprising as much of the literature sites
initial time as the time spent planning and utilizing WM to create and implement problem solutions. Future research could be conducted to measure when WM is used during the solution process and whether pre-planning or online planning is more common when completing the TOL. Additional consideration must be given to the validity of the nBack as a measure of WM. Given the weak relationship between WM and the TOLDX, a commonly used measure of WM, a future study could employ a different measure of WM to more effectively assess this relationship.

The Welch’s t-tests revealed a significant difference between nearly all measures of performance on the manual and computerized TOLDX across all of the age groups. The only exception being execution time in participants in their 20s and older. Execution time being similar between the two tests is indicative of the initial time and completion time increasing proportionally to one another between the two versions of the test. Looking solely at the execution time, it would appear that participants do not solve the computerized test differently than the manual version, just slower.

Although a smaller discrepancy was desired, the results were not unexpected. Gates and Kochan (2015) discussed how creating a computerized test was not always a simple process of taking an existing test and placing it on a computer interface. This method will usually result in some component of the test changing so that the two versions are no longer measuring the same construct(s). The present study found this to be the case in the varying scores of participants between the manual and computerized TOLDX. However, Gates and Kochan (2015) also stated that when scores on the two tests are not highly correlated, the next step is to create a new normative base and psychometric analyses. The present study followed these steps and introduced lifespan
norms for a computerized version of the TOL\textsubscript{DX} using the age groups provided in the original Culbertson and Zillmer (2001) study. This can serve as a basis from which to address the differences that were discovered and to establish the psychometric properties of a computerized version of a commonly used neuropsychological test.

**Limitations and Future Considerations**

One of the major limitations of the present study was the lack of standardization in which the participants completed the TOL. As mentioned previously, the research lab has three different versions of the TOL; one manual and two computerized. Although the date of completion was recorded upon completion of the test and no participant was allowed to complete more than one version in any given session, they were not given the tests in any particular order across numerous testing sessions. As Berg et al. (2010) discussed, familiarity is one of the factors that can have an effect on TOL performance. As the participant is exposed to more versions of the TOL, they could potentially begin outperforming other members of their cohort, above and beyond their actual level of planning ability. This risk is particularly high in regards to the two versions of the TOL\textsubscript{DX} since the problem sets and solutions are identical. It is recommended that future studies of this nature follow a protocol in order to ensure further standardization of not only the test itself, but the order in which they are completed.

Another limitation was the failure to track how the participants completed the TOL\textsubscript{DX}. All participants completed the TOL\textsubscript{DX} on a desktop computer and some of the monitors were equipped with touchscreen capabilities. Participants were not screened after completion as to whether they utilized the mouse or touchscreen. However, anecdotal evidence indicates that a vast majority of participants utilized a mouse when
completing the test. While it can be assumed that many younger individuals are more familiar with touchscreen interfaces, this same proficiency is less likely to be attributed to older adults. It is unknown whether or not the means of completion or computer literacy added additional variance in participants with less computer experience. This could be a possible area for future investigation and studies could indicate whether or not there is a meaningful difference between touch screen and mouse completions. It was also suggested that regardless of familiarity with computers, all participants could complete a brief computer literacy training through tasks unrelated to the TOL or computerized test at hand.

While the TOL\textsubscript{DX} remains a commonly used version of the TOL, it is important to note that modern researchers have made advancements to TOL test creation and analysis. Berg, Byrd, McNamara, and Chase (2010) introduced the idea that certain aspects of the TOL problem structure could allow for greater control over TOL difficulty. This allows researchers to more effectively control how greatly, and incrementally the difficulty will increase over the length of the test. While a brief summary has been provided here, a full description falls outside the scope of this experiment.

The present study sought to create and assess a computerized version of a well-established manual test of planning ability, the TOLDX, and the effect that age would have on performance. Although the present study found a significant difference between performance on the manual and computerized versions, a new normative sample was created to begin establishing the basis for the use of a computerized test. Additionally, the criterion validity was assessed by investigating factors that had commonly affected manual TOLDX performance; age, fluid intelligence, visuospatial ability, and working
memory. Although these results were somewhat inconsistent with past research, the effects were all in the correct direction and supported the utility of a computerized version of the TOLDX. With the improvements being made to, and the increasing use of technology in the field of psychology, it is important to begin assessing the utility and validity of computerized measures and assessments. Given the rigors of test creation and adaptation, the earlier, the better.
REFERENCES


## APPENDIX A

*Correlation Analyses for Outcome and Predictor Variables*

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Pearson-r</th>
<th>Age</th>
<th>FI</th>
<th>Visualspatial Ability</th>
<th>WM</th>
<th>Education</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excess Moves</td>
<td>.10**</td>
<td>-.36**</td>
<td>.38**</td>
<td>-.25**</td>
<td>-.05</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>648</td>
<td>168</td>
<td>639</td>
<td>141</td>
<td>571</td>
<td>43</td>
</tr>
<tr>
<td>Initial Time</td>
<td>.17**</td>
<td>.08</td>
<td>-.03</td>
<td>.02</td>
<td>.06</td>
<td>-.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>646</td>
<td>168</td>
<td>141</td>
<td>153</td>
<td>570</td>
<td>637</td>
</tr>
<tr>
<td>Execution Time</td>
<td>.45**</td>
<td>.56</td>
<td>-.45**</td>
<td>-.24**</td>
<td>.05</td>
<td>.14**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>638</td>
<td>166</td>
<td>138</td>
<td>151</td>
<td>563</td>
<td>637</td>
</tr>
<tr>
<td>Completion Time</td>
<td>.46**</td>
<td>-.47</td>
<td>-.43**</td>
<td>-.20*</td>
<td>.07</td>
<td>.08*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>639</td>
<td>166</td>
<td>139</td>
<td>151</td>
<td>563</td>
<td>640</td>
</tr>
</tbody>
</table>

*Note.* 1 = Male. 2 = Female

**p < .01. * p < .05.
APPENDIX B

IRB APPROVAL

University of Colorado Colorado Springs
Institutional Review Board (IRB)

REQUEST FOR CONTINUING REVIEW (RENEWAL) FORM

Please note: IRB Training is required for all PIs. Faculty Advisors and Personnel involved
with human subject research and must be completed PRIOR TO CONTINUING REVIEW. If
you do not provide the Completion Report Number (located at the bottom of the Completion
Report) and date of your most recent training, your protocol will be returned
to you without IRB review. Go to www.citiprogram.org and follow the instructions
to complete the training.

A. Are you still collecting data on new or previously recruited subjects? ☐ Yes ☐ No
   Are you still analyzing data? ☐ Yes ☐ No
   (If you answered "Yes" to either of the above questions, complete the form and
   submit for continuing review. If you answered "No" send an email to irb@uccs.edu
   with your protocol title and IRB number with a request to close your protocol.)
   (If collection and analysis is complete, the protocol may already be closed. If uncertain,
   contact the IRB.)

B. PRINCIPAL INVESTIGATOR
   Name: Hasher P. Davis
   IRB Training Completion Number: 1060219 Most recent IRB Training Date: 07/15/2014
   Check one: ☐ Faculty/Staff □ Current UCCS Student
   Department, Center, or Institute: Psychology
   Mailing Address: Department of Psychology, University of Colorado, Colorado Springs, CO
   80918
   Phone: (719)255-4148
   UCCS E-mail address: hsdavis@uccs.edu

C. CO-PI (IF APPLICABLE)
   Name:
   IRB Training Completion Number: ____ Most recent IRB Training Date: ____
   Check one: ☐ Faculty/Staff □ Current UCCS Student
   □ Non-UCCS Personnel (Note: Non-UCCS personnel must be approved).
   If checked, explain the role of the non-UCCS personnel: ____
   Department, Center, or Institute: ____
   Mailing Address: ____
   Phone: ____
   UCCS E-mail address: ____

D. FACULTY ADVISOR (REQUERIED FOR STUDENTS)
   Name: ____
   IRB Training Completion Number: ____ Most recent IRB Training Date: ____
   Department, Center, or Institute: ____

Version 3.2.2016