A Tale of Two Towers: The Impact of Problem Difficulty on Task Equivalence Among Preschool Children

by

John Christopher Sheehan
B.S., University of Utah, 2011

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Psychology

Supervisory Committee

A Tale of Two Towers: The Impact of Problem Difficulty on Task Equivalence Among Preschool Children

by

John Sheehan
B.S., University of Utah, 2011

Supervisory Committee

Dr. Kimberly Kerns (Department of Psychology)
Supervisor

Dr. Ulrich Mueller (Department of Psychology)
Departmental Member

Abstract

Supervisory Committee
Dr. Kimberly Kerns (Psychology Department)
Supervisor
Dr. Ulrich Mueller (Psychology Department)
Departmental Member

The Towers of Hanoi (ToH) and London (ToL), historically held as isomorphic measures of problem solving, have had their equivalence questioned in recent years. Adult studies that equalized administration and task structure have found increased correlation between Towers, but insufficient evidence exists regarding whether the same results would be found in young children. This study examined 29 typically developing preschoolers. Both Towers, along with four measures of executive function, were administered in two sessions. The Towers were strongly correlated, but the strength of this relationship was impacted by discontinuation type (i.e., quit vs. standardized), and analyses revealed differences in sustained attention and Tower correlations for those who quit. Complex Tower items showed stronger correlations, and Tower performance and visuospatial WM were also highly correlated. Overall, these results suggest that the Towers, when equated in administrative and structural features, are interchangeable measures of problem solving in preschoolers.

Table of Contents

List of Tables

List of Figures

Dedication

To my partner, my rock, and an endless source of love and support. Thank you, Myrica.

1- Introduction

Executive Function (EF) is often considered to be an umbrella term for those cognitive processes that are fundamental to the ability to react to novel situations in an adaptive and goal-directed manner (Hughes & Graham, 2002; Lezak, 1995). EFs emerge from various interactions between cognitive and emotional processes, resulting in volitional, planned, decisive, and purposive action that efficiently guides behaviour (Luria, 1976; Miyake, Friedman, Emerson, Witzki, & Wager, 2000). Core components of executive function (EF) develop during the infancy and preschool periods that are fundamental to the development of higher-order cognitive processes that will continue to develop into adulthood (Garon, Bryson, & Smith, 2008). Research in EF has consistently shown a gradual improvement in performance that supports the notion of continual, protracted development throughout the childhood years, though this maturation may occur in a number of stages or spurts (Anderson et al., 1996; Bull, Espy, & Senn, 2004). With interest in EF rapidly increasing in recent years, Zelazo & Müller (2011) pointed out several key developmental findings on the topic that have emerged, some of which include 1) the early emergence of EF in development, likely around the end of the first year of life; 2) that important developmental changes occur between the ages of 2 to 5 and continue through adulthood; and 3) that various situations at different ages produce EF failures that can be attributed to the complexity of inferences required.

Research has consistently demonstrated that EF is an elusive construct for measurement (Miyake & Friedman, 2012) and that there are specific circumstances associated with EF assessment in preschool children that are especially troublesome.

Historically, one of the many obstacles to EF research among young children was a general lack of age-appropriate tasks (Garon et al., 2008). In recent years there has been a spike in preschool EF research, as new tasks, or modifications of previous adult versions, have been made available for younger populations (Carlson, 2005; 2012). However, adult EF is theorized to be structurally and functionally different from child EF, particularly compared to preschool children, so tasks designed for use with adults may not always be appropriate for use with children (Zelazo & Müller, 2011). Thus, investigation into the use of common EF measures with young children, as well as into those that have been adapted for use with this population, is important. The emergence of problem solving and planning abilities, which is thought to take place during the preschool years (Bull et al., 2004), is widely regarded as important developmental achievements.

The development of planning and problem solving abilities is hypothesized to include EF processes such as the inhibition of non-relevant stimuli and working memory capacity for maintaining goal representations (McCormick & Atance, 2011; Uneterrainer & Owen, 2006). One popular game used to measure problem solving and planning abilities is the Tower task, which is based on a popular puzzle format that has 3 pegs with a number of disks or balls that can be stacked in different configurations. The task requires that participants move the disks/balls from a ‘starting position’ to a ‘goal position’ on the pegs in as few moves as possible (a ‘move’ typically being counted as when an individual relocates a disk/ball from one peg to another and completely removes their hand from it). Movements are constrained by some basic rules (e.g., move only one object at a time, do not put a larger object on top of a smaller one), which are thought to force the participant into planning out how they will solve the puzzle within the number of moves allowed. Tower tasks have a number of different items, with variations on the start and goal states, as well as in the number of moves that are required to solve the puzzle. The Tower of Hanoi (ToH; Simon, 1975) and Tower of London (ToL; Shallice, 1982) tasks are two traditional versions of this game, having been originally designed for use with adults suspected of having EF deficits. Though the two tasks are very similar in nature, they differ in the sizes of both the pegs and the disks/balls. The ToH utilizes equally tall pegs with disks that are graduated in size, while the ToL uses three equally sized balls, but has graduated pegs. In other words, it is the size of the disks that restricts movement (i.e., via the rules previously described) in the ToH, while movements in the ToL are constrained by the different peg heights.

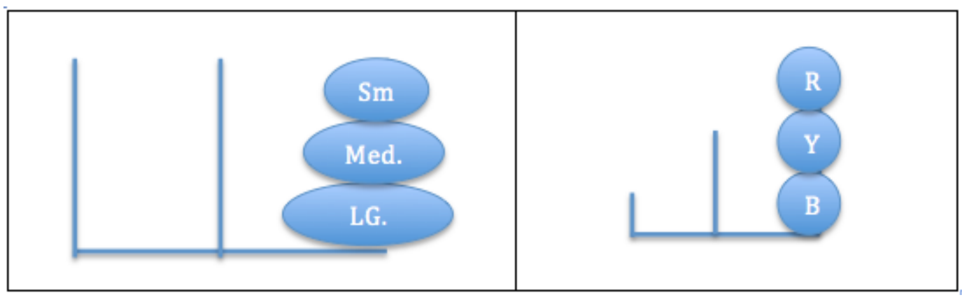


Figure 1- Traditional Tower Formats

Example of what the traditional ToH (left) and ToL (right) look like. The ToH, with equally sized pegs, is restricted in movement by graduated disks that are required to only be stacked atop larger ones. Alternatively, the ToL restricts movement via three graduated pegs, which can hold one, two, or three equally sized coloured balls (red, yellow, and blue). In both tasks, the disks or balls must be moved between the pegs from a start state to a goal state in as few moves as possible.

A long-held view of the two tasks was that they are isomorphic and tapping into the same underlying cognitive processes (Unterrainer, 2005). However, many studies that have attempted to correlate performance for the tasks have been unable to find a consistent level of correlation, if any, thus calling into question whether these two tasks tap into the same executive processes (e.g., inhibition, working memory, etc.; Humes, Welsh, Retzlaff, & Cookson, 1997; Schnirman, Welsh, & Retzlaff, 1998; Welsh, Satterlee-Cartmell, & Stine, 1999; see below for discussion of these studies). This study seeks to gain a better understanding of the administrative and structural differences between the ToH and ToL and how these differences affect the performance of preschool children.

1.1- Tower Tasks

Classical disk-transfer tasks, specifically the Tower of Hanoi (ToH; Simon, 1975) and Tower of London (ToL; Shallice, 1982), have been extensively used as putative measures of executive functioning in adults and, to a lesser extent, child populations (Bull et al., 2004; Unterrainer, 2004; Welsh, Satterlee-Cartmell, & Stine, 1999). Studies document the goal-directed behavior, planning abilities (to maintain rules), formation of subgoals (for advancing to the final goal), and execution of movements required by these challenging tasks (Anderson & Douglass, 2001; Goel & Grafman, 1995;Salnaitis, Baker, Holland, & Welsh, 2011; Simon, 1975).

Performance places demands upon working memory resources, as these various elements must be maintained while in pursuit of the final goal (Salnaitis et al., 2011). Shallice (1982) introduced and argued that the ToL task was a particularly useful measure of planning and problem solving, given its limited dependence upon lower-order cognitive skills (e.g., spatial processing, visuo-motor coordination, etc.) and ability to analyze executive skills (e.g., planning speed, impulsivity, and flexibility; Anderson et al., 1996). He noted the ToL to be a measure of planning that provided a greater variety of qualitatively different problems than the ToH, due to structural differences.

1.2-Tower Tasks for Young Children

Despite being originally designed for use with adults, there are a number of components of Tower tasks (e.g., short administration time, game-like presentation, etc.) that make them particularly well suited for the limited attentional resources of young children. One notable advantage of Tower task use with children versus older populations is the ability of children to be able to complete simple problems at a relatively early age (e.g., 3-years-old), but not being impacted by ceiling effects (successful completion of all problems) that impacts the range of scores for older populations (Welsh et al., 2000). However, some of the first studies using the ToH with young children suggested that 5- and 6-year-olds have great difficulties with the task, as they were seen as only being able to solve the 2-disk set with great effort and the 3-disk (or more) sets were almost unachievable (Byrnes & Spitz, 1977; Piaget, 1976). Klahr and Robinson (1981), perplexed by the limited ToH success in young children, observed that 5- and 6-year-olds are very often able to demonstrate problem-solving activities in the real world that required multiple planned steps. These real-world activities would, in their view, suggest that young children do possess the capacity to plan beyond the removal of, at least, a single obstacle en route to a larger goal.

1.3- ToL & ToH Comparison Studies

Generally speaking, there are two common methods by which the ToL and ToH have been compared. One approach is a between-groups comparison, administering each Tower task to a separate randomized sample, and using identical criterion measures of various EFs in both groups (e.g., Bull et al., 2004). Those who utilize this approach often cite the ability to eliminate any practice or carry-over effects that would occur in a within-subjects design Tower. The comparison of the two tasks is dependent upon how the specific criterion outcomes are measured in each Tower type (Bull et al., 2004). The other common approach is a within-subjects design, administering both Tower tasks to the same participants dependent on how specific criterion outcomes are measured (e.g., Unterrainer et al., 2005; Welsh et al., 1999; Zook, DeLosh, & Davis, 2004). Utilizing this approach, Welsh et al. (1999) reported only a modest correlation between total scores on the ToH and a 25-item version of the ToL (i.e., ToL-R) of .37. The finding of modest correlation was also noted by Humes and colleagues (1997). Welsh et al. (1999) notes these findings closely resemble previous work on the intercorrelations of the ToL-R and ToH, with a range of correlations falling between .39 and .60.

Studies utilizing the ToL-R with 30-items reach the higher end of this range, with Satterlee-Cartmell (1997) finding a correlation of .60, though some studies find very low correlations, such as Zook and colleagues (2004), who report a correlation of .27 (unable to account for 93% of the variance between the two tasks). Even studies with higher correlations reaching the upper-limits of the range proposed by Welsh et al. (1999) find a substantial amount of the variance between two tasks is unaccounted for. Indeed, various studies comparing the ToL/ToL-R and the ToH have reported correlations that leave approximately 75% or more of the variance in scores unshared between the two tasks (Humes et al., 1997; Welsh et al., 1999; Zook et al., 2004). Welsh, Revilla, Strongin, and Kepler (2000) sought to determine whether the non-shared variance between the two Tower tasks was attributable to administrative differences. They adjusted for basic administrative differences, such as the number of trials per item/problems, but did not address differences in item/problem difficulty (see section below on 'problem difficulty') beyond matching the items on the two Towers for minimum move count. Their findings showed a similar correlation (r= .66) as previous work, leading them to conclude that administrative differences were negligible.

1.4- Differences in Tower Tasks Used in Research

Most neuropsychological research using Tower tasks has been based on use of, or adaptation to, the original ToH or ToL tasks. The overall similar goals and overlap of the physical aspects between the ToH and ToL contributes to an assumption of inter-changeability between the two tasks. They both, at their most basic level, involve the act of transferring objects from a start state to a goal state on three pegs in a proposed minimum number of moves. The solution is guided by a handful of similar rules in both tasks, such as moving only one object at a time while others remain stationary on the pegs. However, beyond these surface similarities there are many fundamentally different structural components that researchers cite as possible factors to substantial non-shared variance that exists between the two tasks (Kaller, Unterrainer, Rahm, & Halsband, 2004; Salnaitis, Baker, Holland, & Welsh, 2011). Additionally, studies often modify the traditional tasks, which only adds to the number of fundamental differences between the two Tower tasks.

1.5- Physical Structure

Differences between the two Towers include how the physical characteristics and game rules constrain the potential moves an individual can make. In the TOH, a game rule limits how the player can place disks on the three equally sized pegs (i.e., a smaller disk can be placed on larger ones, but never vice-versa). On the TOL, however, the pegs are graduated in size, limiting the placement of three differently colored but equally sized balls by how many can fit on each peg (one, two, or three per peg). Shallice (1982) believed that the alterations in the ToL task allowed for more complex problem sets (which ranges from 2- to 5-move solutions on both Tower versions) than were possible in the ToH. However, various alterations to the original structure have been made in order to extend the use of the Tower task to different populations. Variations have either sought to create more difficult tasks than the original versions for use with healthy adult populations, or have attempted to simplify the task for use with children (Unterrainer et al., 2005).

1.6- Modifications to Basic Task Demands

In an attempt to create a more difficult task, Ward and Allport (1997) increased the number of ToL balls to five and equalized the length of the three rods such that each was capable of holding all five objects. This alteration allowed for planning sequences up to 13 moves in length, a substantial increase from the original version. Kaller et al. (2008), however, noted that this physical structure change may be to the advantage of young children due to the elimination of the constraint on how many balls can be placed on each of the graduated pegs. Kafer and Hunter (1997) also created a version with the same goal of added difficulty that increased the number of balls and rods to four each.

1.7- Modality

Another important area of consideration when comparing Tower tasks is the modality of Tower presentation. Indications that there might be variability in Tower performance based upon differences in modality arose very early on in Tower studies, as stark differences were found between young children's performance on the traditional ToH (Byrnes & Spitz, 1977; Piaget, 1976) and a modified, 'mental' version (i.e., no motor component) of the ToH that had been designed by Klahr and Robinson (1981). Recent variations in Tower modality have also emphasized the use of computerized versions of the Tower task. Mataix-Cols and Bartres-Faz (2002) used both a wooden and a computerized version of the ToH to assess the impact that modality has upon performance. They found no significant differences in a study with undergraduate college students on any of the dependent variables (e.g., number of moves, errors, reversions, and time to solution; but see Noyes & Garland, 2003; Salnaitis et al., 2011; and Williams & Noyes, 2007 for further discussion on the subject). Interestingly, they also noted a lack of transfer effects from one task to the other, even though the two Tower tasks were administered sequentially. This contradicted previous finding by Adams et al. (2003), which found a learning effect when a computerized Tower task was administered prior to the manual task, but not vice-versa. With regard to the use of computerized versions with preschool populations, Kaller et al. (2008) found that, with some practice items, children four years and older, without previous computer experience, were able to effectively utilize a touchscreen to complete Tower tasks. The investigators chose a touchscreen given previous reports of potential confounds for planning and motor skills seen in child studies when using the traditional computer mouse (Luciana & Nelson, 1998).

1.8- Instructions

In addition to the physical differences between the two tasks, differences exist from the very beginning of administration. The directions that the participants receive at the beginning of the tasks differ for the two traditional Tower versions, not to mention among the various alternative versions of each task. For example, in the classic ToH instruction, there are no specifications as to the number of moves that are required in order to achieve success. In the ToL task, however, participants are instructed to complete the task in a specific number of moves as *quickly* as possible (Welsh et al., 1999). Bull et al. (2004) raised the question of whether instructional differences, such as hearing a rule of doing this quickly, might lead participants to begin moving the balls before they engage in planning processes (i.e., contributing to impulsive initial moves). They suggest that this variation in task instruction may lead to variations in strategy development or use and could impact reliance upon planning and other underlying cognitive functions. Kaller et al. (2008) attempted to eliminate this problem using a modified version of the ToL, removing any suggestion of minimum moves and further having the computer program acoustically prompt children before each item to *plan ahead* first. This instruction prompt, used previously has been found to encourage initial planning over 'on-line' or 'in-the-moment' planning (Davies 2005; Unterrainer, Rahm, Leonhart, Ruff, & Halsband, 2003). Klahr and Robinson (1981) suggested that task modifications often made for children, such as a monkey-themed cover story, interesting objects, and other environmental elements, can assist in sustaining the attention of children so they are better able to make an effortful attempt at solving multiple problems. Fundamental administrative and structural differences between the two Tower tasks, as well as differences arising from variations on the traditional measures, are discussed below.

1.9- Errors

The manner in which errors are defined, measured, and dealt with are all sources of variation in scoring Tower tasks. Bishop and colleagues (2001), required the tester for all errors to remind the subject of the rule, restart the trial, and counted any trials with an error as failed. In this study, they also charged children with two moves if they moved a disk to another peg and back, while maintaining their hold on the disk, a substantial deviation from some traditional protocols. Task administration may play an important role in error scoring as well, as some computerized Tower tasks limit the ability to make certain types of errors. For example, the common rule that you may not pick up two disks/balls at the same time is something that would be an inviolable rule in a computerized version. While this type of error could be made and recorded in a physical version of the task, a computerized version prevents this type of error from even being made. The computerized ToH version from Kerns and McInerney (2007) logs attempted illegal moves made by the child, while also informing children of their mistake via a large stop sign that appears on screen. This immediate feedback (i.e., the stop sign) helps children recognize that their intended move was a rule violation; something that might go unnoticed if there had been no indication that the move was an error. Given that the purpose of the task is to assess problem solving, not to determine whether a child can remember the various rules of the task, this stop sign is intended to help to minimize task confusion so that deficits in performance can be attributed to actual problem solving abilities.

1.10- Difficulty Level: Problem Structure & Move Type

In many studies using Tower tasks, mandating a minimum number of moves allowed for successful completion of a item has been relied upon as the primary means to adjust task difficulty. A growing amount of research suggests that a concept of problem difficulty that just includes minimum moves is a poor indicator of performance (Berg & Byrd, 2002; Borys, Spitz, & Dorans, 1982; Kaller, Unterrainer, Rahm, & Halsband, 2004; Ward & Allport, 1997). Various studies have identified additional factors that contribute to difficulty level, including factors related to the general nature of the problem/item (also called global task problems) and those that are related to each individual move (local task problems). Studies that have attempted to equalize both types of problems across both Tower versions have found that participants have similar performance (e.g., final scores) even when the physical structures of the tasks (i.e., task versions, ToH vs. ToL) are different (Unterrainer et al., 2005). The authors have interpreted these findings as suggesting individual problem complexity/difficulty is a more significant indicator of Tower performance than the nature of the physical structure itself; which in direct opposition to Shallice's (1982) assertion that it is the structural differences between the two Tower versions that impact performance the most

As noted earlier, a 'move' is generally considered to occur when participants relocate a disk/ball from one peg to another, removing their hand completely from the object in order for the move to be considered finished. The number of minimum moves required for the first item typically ranges from one to three, depending largely upon the population of interest. While easier items (i.e., lower minimum move) are able to be solved via a "perceptual match-to-sample strategy," those items which have a higher number of minimum moves require goal directedness, planning and execution of subgoals (e.g., moving one disk/ball out of the way to gain access to a disk/ball below it), inhibition of alternative responses, and effective working memory to guide actions

(Albert & Steinberg, 2011; Berg & Byrd, 2002). While various adult Tower tasks have had a much higher ceiling (e.g. Shallice's [1982] original version), Tower studies focusing on young children have typically had a maximum range of six to ten moves (Bishop et al., 2001; Bull et al., 2004; Kerns & McInerney, 2007). Minimum number of moves, however, is not the only contributing factor to problem difficulty. As Berg, Byrd, McNamara, & Case (2010) note, other factors, such as start and goal position hierarchy (i.e., whether the disks/balls are all stacked on one peg, all on different pegs, or two on one peg and one on another), subgoal requirements, and the number of solution paths available (i.e., how many different 'routes' can be take to get from the start state to the goal state), are also critical in determining the difficulty of a problem.

Goal hierarchy on Tower tasks relates to the level of ambiguity that exists within the final state of the disk/ball configuration (Kaller et al., 2011). For example, when all of the objects in the final goal state are stacked upon each other on one peg, it is considered to be an unambiguous Tower-ending state (T-ending), as it is clear which disk/ball must be placed into its final position first, second, and so on. A partially ambiguous, "partial Tower-ending state" (PT-ending), however, has two or more objects upon a single peg, but leaves at least one other object elsewhere, thus creating some uncertainty as to the order with which the objects must be placed. The most challenging is the flat-ending state (F-ending; i.e., completely ambiguous), as there is no indication as to any ordering sequences whatsoever (Berg et al., 2010; Kaller et al., 2004). Klahr and Robinson (1981), in their verbal-only Tower task, found dramatic differences between F-ending and T-ending conditions. For example, they found that, while most 5-year-olds and almost all 6-year-olds were able to give perfect 6-move plans on T-ending problems, only about one-third of 5-year-olds could reliably do the 3-move problems in the F-ending condition. Additional differences appear in how the goal-state hierarchies are represented in the traditional tasks, with the ToH utilizing a physical model, as opposed to the pictured end-state goal that has been used in the ToL. Bull et al. (2004) noted that the more salient physical representation could be more advantageous for younger children.

Another consideration in regard to task difficulty is ‘suboptimal solution paths,’ which is a longer sequence of moves (than the minimum required for a given Tower item) to the goal state, which can also run the risk of lengthy detours or dead ends (Kaller et al., 2011). A detour occurs when the participant places a disk/ball onto the goal peg, but at an incorrect time (i.e., before other disks/balls have been placed), requiring that additional moves are made to progress towards the correct goal state. A dead end requires even more additional moves to be made after a goal move has resulted in blocking the subsequent solution of the problem, meaning that moves must be reversed in order to proceed. While an analysis of individual move sequences would need to be made in order to determine the number of detours and dead ends that occurred in a given item, the simple presence of additional moves beyond the minimum number is an indication that a suboptimal path was taken.

Tower tasks are generally thought of as problem-solving measures that require planning activities in order to achieve success. However, actual planning is not always necessary in order to successfully navigate a Tower task. ‘Perceptually driven moves’ occur when the next move that is made by a participant is directly guided by the current object configuration (Bull et al., 2004). In other words, no planning is required, only ‘real-time’ movements based upon the perception of what action will bring the configuration closer to the end-state goal. Perceptual strategies were also discussed by Simon (1975), though a distinction should be made here between a relatively simple and stimulus-driven approach and a more intentional planning strategy that is based upon one's perception of the current Tower state in relation to the goal state. One- and two-move Tower problems, for example, only require these perceptually driven moves to be made, as each move towards the goal-state lands the disk/ball in its final position (Kaller et al., 2008). However, it is important to note that, while one- and two-move problems do not explicitly require planning, this does not entirely justify the assumption that the participant is still not engaging in any planning. In such cases, it is difficult to determine whether or not any actual planning has occurred, though it is possible that there could be differences detected in latency time prior to movements. In their work studying age effects on ToL performance, Albert and Steinberg (2011) noted differences between age gains made on 'easy' problems vs. 'hard' problems, with the former ceasing to improve beyond the age of 17 and the latter seeing gains into the 20s. They noted the major difference being the presence of intermediate subgoals that required planning and execution (as opposed to 'easy' problems).

While perceptual moves do not explicitly require planning activities to occur in order for success, Global-local goal conflict moves (GLGC moves) are unable to be successfully navigated without the inhibition of the perceptual response and subsequent planning and execution of new subgoals (Kaller et al., 2011). GLGCs require the participant to take the object further away from the goal state in the immediate future (i.e., "locally"), in order to accomplish the optimal solution path in the long run (i.e., "globally"; Kaller et al., 2011). A sub-category of GLGCs, counterintuitive moves

(CIMs), require that a participant move a disk/ball that has already been placed on its goal-state (note that the terms GLGC and CIM are often, but not always, used interchangeably). The importance of GLGCs stems from the suggestion made by Goel and Grafman (1995) that item difficulty is actually dependent upon differences in complexity that GLGCs creates. Therefore, as Bull et al. (2004) note, not only do these moves create the need for inhibition and planning, but working memory and shifting may also be required to remember and flexibly shift the subgoals needed to reach the goal-state.

Studies with both adults and children (including preschoolers) have consistently found that problem structure (e.g., depth of search and goal hierarchy) does influence the accuracy of planning in children (Kaller et al., 2008; Klahr & Robinson, 1981). A key difference exists between two- and three-move problems in terms of problem structure, as two move problems (regardless of Tower task) will never require a child to produce a counterintuitive or intermediate move in order to complete the item. Both the ToL and ToH, however, include three-move problems that may require that these more complex moves be made. Numerous studies have observed dramatic drop-offs in performance in younger children (e.g., 3- and 4-year-olds) on three-move problems (Klahr & Robinson, 1981; Luciana & Nelson, 1998; Welsh, 1991), something that Kaller et al. (2008) explain as being a failure of these children to look ahead. In their study of preschoolers' performance on both the ToL and ToH, Bull et al. (2004) found that the role of shifting abilities in both Tower tasks was increasingly prominent on items that required more GLGCs/CIMs for successful completion. That Tower tasks appear to show a higher association when taking item difficulty into account seems to be in line with the work of

Webster and Borys (1982), who noted that subgoal length was a more sensitive index of task mastery. However, despite performance on a measure of inhibition being related to ToL items with two or more GLGCs, the ToH did not show any such relation to that inhibition measure. McCormick and Atance (2011) still suggest that inhibitory control is required for the type of cognitive flexibility that is needed to navigate more complex Tower items, and that it is likely to be related to child performance on Tower tasks. It is still unclear, though, whether the complexity of the items on the ToL and ToH, when other administrative differences are minimized, would significantly impact how strongly the two Towers are correlated in young children's performance.

1.11- Discontinuation

Discontinuation on a given item typically occurs when the participant (1) reaches the goal-state, (2) performs a certain number of moves above the minimum number required for solution, or (3) a time limit is reached. Albert and Steinberg (2011) had a time limit of 160 seconds, but they also required that their participants 'submit' their computerized ToL by clicking a button. So, theoretically, participants could have reached the solution, but not have had the opportunity to hit 'submit' until after time ran out (or, alternatively, simply forgotten this requirement). Various studies have used different criteria for 'maximum' number of moves. While some use a strict and constant number (e.g., 20, as with Bull et al., 2004) as their cutoff, others utilize varying formulas that add the minimum number of moves to a specified number of additional moves (e.g., Kerns & McInerney, 2007). The latter approach seemingly avoids the risk of requiring an increasing amount of efficiency, as the minimum number of moves grows closer to the cutoff (e.g., 20 moves). For example, using a strict 20-move cutoff on a 3-move Tower problem would give the participant an additional 17 moves, which would inevitably increase the likelihood that the goal state could be reached by chance alone. However, using the same cutoff of 20 moves on a 7-move problem would only provide an additional 13 moves, meaning that there is a smaller margin for error in the route the participant takes to reach the goal state. Child Tower tasks have typically been discontinued when two consecutive items are failed, children refuse to continue, or no legal moves are made on a given item (Bishop et al., 2001; Bull et al., 2004; Kerns & McInerney, 2007).

1.12- Scoring

Because the various scoring methods utilized by the ToL and ToH (and their variants) are often dependent upon their unique administrative and structural qualities, as well as the underlying interests of the researchers, scoring methods vary across task versions. For example, Welsh et al. (1991), in their work with the ToH, calculated a *planning efficiency* score based upon the number of attempts it took a participant to produce two consecutive 'optimal solutions' (defined as reaching the goal state in the minimum number of moves). The scoring system gave participants higher scores for fewer attempts needed to produce two consecutive optimal solutions (e.g., two successive solutions on 2^{nd} and 3^{rd} attempts earned more points than optimal solutions on the 3rd and 4^{th} attempts). This type of scoring approach, however, is only appropriate for Tower tasks administrations that administer multiple trials of the same item to some criteria. Kerns & McInerney (2007), using a computerized version of the ToH that provided only one attempt per item, calculated a ratio of the total number of moves across all items to the number of items completed correctly. This ratio provided the average number of moves made per successful item, adjusted for the number of errors (i.e., attempted illegal moves) made.

Summary scores, which represent various aspects of performance beyond the number of items passed, have also been utilized in several recent studies. Anderson et al. (1996) utilized a single summary score that combined a number of factors, including planning efficiency (total number correct), impulsivity (number of failed attempts), and speed of performance. Their standardized scoring system also has the added benefit of providing developmentally referenced information that can allow clinicians to determine how close a score is to what would be expected for a particular age. Albert and Steinberg (2011) used the percentage of trials solved in the minimum number of moves, which they termed "perfect solutions," as their primary variable. This, they suggest, is an indication of effective planning and subsequent execution, and this score was correlated highly enough with non-perfectly performed items (i.e., those requiring 'extra' moves beyond the minimum), that Albert and Steinberg (2011) decided to limit their analyses to perfectly solved items only. This was also a reflection of their interest in the development of strategic planning, rather than more perceptual or online problem solving. This study also found that first-move latency (i.e., time between problem presentation and first move) was strongly predictive of performance and, therefore, was interpreted as an indication that strategic planning plays a major role in TOL success.

Alternatively, Bishop et al. (2001) scored the total number of items successfully completed by a participant. In their study, a final score was given to match the highest level/item achieved (e.g., a item that required 5 minimum moves would be worth 5 points) successfully completed. Furthermore, because two problems were administered per difficulty level, they decided to award an additional half-point if the participant was able to complete both items at that difficulty level. While this strategy captured the highest difficulty level that a participant was able to achieve, this score failed to represent the actual number of successful items to that point. Although, again, it is important to keep in mind what the variables of interest are for a particular study. Seeking to capture both the highest level of complexity achieved, as well as the consistency of success across all items, Bull et al. (2004) revised this scoring system by awarding every item passed (with the exception of the 1 move items) with points that matched the minimum number of moves required (e.g., 3 move problem=3 points; 5 move problem=5 points). While this strategy better captures the consistency of performance across all items administered, it still leaves other aspects of performance unevaluated. For example, a participant who completed a 5-move problem in the minimum number of moves (i.e., highly efficient planning) would receive the same score as the participant who completed that item in eight moves (i.e., less efficient).

1.13- The Role of Basic Cognitive Processes in Tower Task Performance

1.13a- Inhibitory Control

Inhibitory control has long been theorized to play an important role in Tower task performance due to the perceived need of participants to refrain from either breaking task rules (e.g., placing a bigger disc atop a smaller one in the ToH) or from making a 'perceptually-driven' move towards the goal state when an alternative move is more appropriate (e.g., using a subgoal). While most researchers agree that inhibitory control is important in Tower tasks, research findings document considerable variation on their association with Tower tasks performance in both adult and child populations. In a study with school age children and adolescents, Bishop, Aamodt-Leaper, Creswell, McGurk, and Skuse (2001) found inhibition to be unrelated to ToH performance. Attempting to explain this lack of association, Bishop et al. (2001) suggested that the conscious inhibition required for moving from one subgoal to another (i.e., shifting) could be a better predictor of ToH performance than a Stroop-like measure of inhibition of prepotent responses. They distinguish between the notion of unconscious/automatically-activated suppression of prepotent information versus the conscious inhibition of previously used task sets or strategies (as might be measured by shifting abilities).

Variation in the correlations between various inhibition tasks and the ToL has also been previously reported. Albert and Steinberg (2011) utilized the Stroop task as a measure of interference control, but then used a latency time variable (i.e., the amount of time that elapses prior to attempting to solve the problem) as their measure of impulse control. They found that the Stroop (i.e., interference control) variable was unrelated to ToL performance, and reasoned, given that there is no ‘training’ built into the Tower protocol that would establish a prepotent response, that this finding was not surprising. They did, however, find that age-related gains in the performance-based impulse control variable was highly related to ToL performance, similar to findings by others (Albert & Steinberg, 2011; Asato et al. 2006). Asato and colleagues suggested age-related increases (e.g., 8- to 13-year-olds vs. 14- to 17-year-olds) in thinking time decreased the likelihood of impulsively attempt items without pre-planning, which was associated with better ToL performance. This suggests inhibitory control is important in planning tasks (e.g., ToL) as it relates to participants taking time to plan/think rather than impulsively responding.

Goel and Grafman (1995) noted specific circumstances in which consistent perceptually driven moves create a prepotent response, which then has to be inhibited as the task grows more complex and a non-perceptual responding (i.e., planning) is required. They suggested that it was not inability to plan that prevented the participants from completing complex problems, but rather an inability to inhibit the strong prepotent tendency to make a perceptual response. Baughman and Cooper (2007) built upon this research, arguing that developmental gains are responsible for inhibition of automatic perceptual responses and a subsequent re-evaluation of the specific ball configuration at that moment (see below). Other studies focusing on child performance on inhibition-tasks and Tower tasks have also found varying levels of association (Asato et al., 2006; Hughes et al., 2010; Kaller et al., 2008; Lehto et al., 2003). However, as Zelazo & Müller (2011) note, inhibition does not sufficiently explain all that is required for successful performance on Tower tasks.

1.13b- Working Memory

Among the cognitive processes that contribute to Tower task performance, working memory is considered to be one of the most important (Bull et al., 2004; Welsh et al., 2000). Previous research in adults has shown visuospatial working memory, in particular, contributes to Tower performance, suggesting adults utilize a strategy in which they mentally visualize the movements when planning a response (Welsh, Cicerello, Cuneo, & Brennan, 1995). However, Bull et al. (2004) argue there is no evidence suggesting that preschool-aged children implement a similar visuospatial strategy. Furthermore, children don't appear to pause prior to responding, indicative of a lack of formulation and implementation of a planned sequence (i.e., in Bull and colleagues view, working memory requirements appear negligible, given their perception that no plan formulation could be occurring). Indeed, this lack of pause before response is a further indication that young children are more prone to engage in an on-line, perceptual approach to Tower tasks. Bull et al. (2004) argue that, due to this tendency, visuospatial working memory does not play a particularly important role in performance, especially when there are either no, or very few, subgoals required for completion (e.g., 1- or 2-move sequence items; Bull et al., 2004; Goel et al., 2001; Shum et al., 2000). One concern, however, is that Bull and colleagues based their proposal on casual observations of a lack of latency time, which has been used by some (e.g., Albert & Steinberg, 2011; Asato et al., 2006) as a measure of impulse control. However, a lack of latency time, particularly when based off of casual observation, is somewhat ambiguous in terms of how working memory in related to Tower tasks, or problem solving in general. Therefore, whether or not working memory independently contributes to Tower performance in preschool children has not been thoroughly investigated.

The reported correlations between performance on the Tower tasks and working memory have varied across studies. Welsh et al. (1999) found no such association between WM variables and ToH performance. Roberts and Pennington (1996) distinguished between those WM tasks that tap into capacity (e.g., span tasks) and those that require online processing (e.g., N-back tasks, a working memory task where subjects are asked to monitor the identity or location of a series of verbal or nonverbal stimuli and to indicate when the currently presented stimulus is the same as the one presented N trials ago). Welsh and colleagues (1999) argue that, while a theoretical separation could aide in understanding the varying levels of association between different WM-specific and Tower tasks, the different WM processes still take place within the overarching system (i.e., a sufficient capacity must exist in order for computations to be made) and therefore should be related. Adding to this unity and diversity discussion of WM, it is important to note that individual Tower items could potentially require varying levels of each type of WM process. For example, a three-move problem on the ToH/ToL may require minimal capacity compared to a six-move problem. Furthermore, one might expect differential computational WM demands between a 3-move problem with no global-local goal conflict moves (GLGC; see sections below) and a 3-move problem with one GLGC move. Thus, it may be important to not only look at how a WM task relates to Tower performance overall, but how it relates to performance on specific types of problems (i.e., those with higher levels of difficulty).

1.13c- Shifting/Flexibility

Though a limited amount of research has focused on the impact of cognitive flexibility and shifting on Tower tasks, there are a number of studies that have assessed shifting abilities in preschoolers via alternative tasks. These studies are helpful to gauge what one might expect the contributions of shifting to be on Tower tasks. In their work assessing task switching in preschoolers using the Dimensional Change Card Sort (DCCS) task, Diamond, Carlson, and Beck (2005) found a developmental progression that occurred between 2½ -year-olds and 3½ -year-olds. They suggested that preschoolers have difficulty considering objects in a manner that differs from their original relevant attributes. In other words, younger preschoolers could have trouble shifting between different types of moves (see below) and therefore be unable to succeed on more difficult items, though would progressively develop this ability throughout their preschool years. In their work, Chevalier et al. (2012) note the contributions of inhibition and working memory to the goal representation component of flexibility (i.e., shifting) by late preschool years, but cite a lack of correlation between these two EFs to a switch implementation component. This is identical to what Brocki and Tillman (2014) reported, as they noted that WM and inhibition are the driving force behind set shifting, specifically goal representation. Placing a greater emphasis on inhibition, specifically action inhibition and attentional inhibition on the DCCS task, Rennie, Bull, & Diamond (2004) suggested that inadequate inhibition may actually be sufficient to account for all errors seen in preschooler performance on this task. Davidson, Amso, Anderson, and Diamond (2006), in their study of memory, inhibition, and task switching, made an interesting observation that even four-year-olds were able to hold information in mind and inhibit a dominant response. This was dependent, however, upon the inhibition being steady-state (i.e., single task blocks) and the rules remaining constant. This could suggest that varying types of moves required within the Tower tasks could alter the state of the problem enough to disrupt these abilities.

1.14- Rationale

Clearly, Tower tasks have utility for understanding aspects of problem solving and as measures of EF in children. Given the scarcity and variability of findings for Tower tasks in preschool children, and the differences between the ToH and ToL, further investigation is warranted to evaluate these tasks, and the relation between overall performance and item complexity (e.g., GLGC moves). It is not clear in previous studies whether administrative and structural differences confounded results, limiting the correlation between the two Tower tasks (Bull et al., 2004; Welsh et al., 1999). Differences in associations between basic cognitive processes, including inhibition, working memory, and shifting tasks, and Tower task performance are also possibly attributable to differences in the design of the two Tower tasks. It is unknown, however, to what extent these differences in correlations are attributable to administrative (e.g., modality, instruction, etc.), structural (e.g., graduated disks vs. graduated pegs), or cognitive complexity differences. While previous studies have attempted to equate the two tasks in a variety of ways, few have attempted to isolate structural differences by creating closely matched structural and administrative features to allow for increased comparability (Unterrainer, 2005). It is also paramount that the individual items from either task be as similar as possible in terms of complexity and difficulty level, to both assess shared variance in overall performance and to isolate reliance upon individual underlying executive functions. The goals of this study were to 1) determine whether matching the two Tower tasks in both structural and administrative features would increase the correlation of the two tasks significantly beyond the previous studies (i.e., outside of the .4-.6 range cited by Welsh et al., 2000), as well as increase the consistency of their correlation with other EF measures, 2) determine whether complex Tower problems (i.e., those with 2- or 3-GLGC moves) are more highly correlated between the two Tower versions than are simple problems (i.e., those with 0- or 1-GLGC move).

2- Methods

2.1- Participants

This study, approved by the Human Ethics Review Board (HREB) of the University of Victoria, recruited fifty-four preschool children between the ages of 3- and 5-years-old from six daycares and preschools in the Greater Victoria area. Parents received recruitment flyers and those interested signed a consent form and returned it to the school for pick-up. The investigator contacted those parents who consented and a phone interview was conducted to screen potential participants for exclusion criteria (see Appendix C) including any neurodevelopmental concerns, diagnoses, or history that included a loss of consciousness. Thirteen children were excluded from participation in the study due to their neurodevelopmental history. An additional twelve children were not able to be included in the sample due to unavailability during testing times (N=7), having limited English language skills (i.e., non first-language speakers; N=2) or the child declining the assent process during one of the sessions (3). The final sample size of 29 included 15 females (average age=55.7 months, SD=5.7 months) and 14 males (average age= 55.2 months; SD= 5.8 months).

2.2- Procedure

Participants were tested individually on two separate one-on-one testing sessions (approximately 10 days apart) in a quiet space at their individual daycare/preschool facility. All testing sessions were conducted by either the male primary investigator or a female research assistant (N=2 participants tested by female research assistant) and lasted approximately (25-30 minutes). The following fixed task order of administration was used, with the two Tower tasks being counterbalanced; Session 1: Tower Task A

(ToM/ToL), Stroop, Go/No-Go, and CPT; Session 2: Tower Task B (ToM/ToL), Jack's Boxes, and FIST. It is important to note that not all children completed all tasks; the specific number children who did not have complete data for each measure is outlined in Figure 2- Missing Data (below). Overall, children most often quit tasks due to losing interest/ becoming bored. Less commonly, a small number children quit the tasks when they either saw their classes doing an appealing activity (e.g., crafts or recess), or, on one occasion, a parent arrived early for pick-up from daycare and testing was unable to resume at a later time due to unavailability.

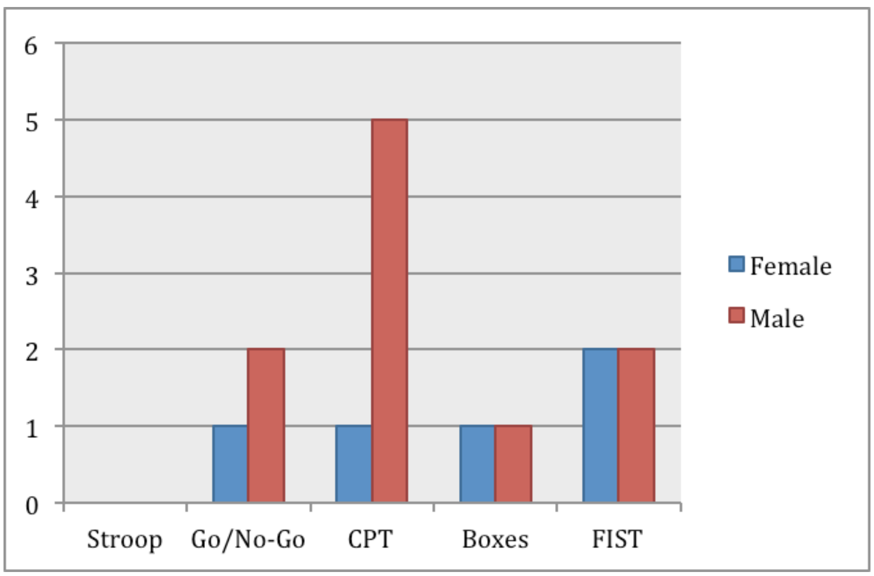


Figure 2- Missing Data

Number of incomplete/invalid tasks per measure, by gender

2.3- Measures

2.3a- Tower of Hanoi (ToH) – Modified Tower of Monkeys

The Kerns and McInerney (2007) Tower of Monkeys (ToM) task is a touch screen computerized version of the ToH that is specifically designed for use with young children. Each child was instructed to transfer three monkeys, graduated in size, from a starting state to a goal state while complying with two specific rules: (1) A larger monkey cannot be placed atop a smaller monkey, and (2) only one monkey can be moved at a time. If at any point a child attempted to break these rules, a large “Stop” sign appeared, an error was recorded, and the investigator reminded the child of the rule that he or she violated. All children were asked to give the administrator examples of legal and illegal moves prior to beginning the practice items. Upon completion of instructions, children were presented with three equally sized trees and three monkeys of graduated size on the screen. In order to reduce the abstractness of the rules and in line with Welsh et al.’s (1991) administrative procedures, a cover story was introduced concerning this family of monkeys (e.g., “Daddy, Mommy, and Baby” monkey to represent large, medium, and small monkeys, respectively) and the trees.

Figure 3- Tower of Monkeys Screenshot

Tower of Monkeys (ToM) version of the traditional Tower of Hanoi (ToH) task. A ‘Daddy, Mommy, and Baby’ monkey can all be seen on three trees of equal size. The vine at the top of the screen is used for the functional purpose of moving the monkeys (i.e., touch the monkey, it moves to the vine, then touch the tree where it is to be placed; this represents a single “move” in the game).

Children were told that the family of monkeys wanted to reach the rightmost tree in order to reach the bananas (placed at the base of this tree on screen). The screen contained a picture of this goal-state (a smaller portion of screen) and a start state (majority of screen space) that was then manipulated by the child to reach the goal. After two practice items (one each of a one- and two-move item), a "start" button appeared for the child to touch, to initiate the first item. Children were not told to complete the items in a certain number of moves, nor to complete the task quickly, but instead were encouraged to try to think where each monkey should go to arrive at the tree with bananas in the right order. Timing was commenced immediately upon presentation of the first scored item to calculate the total time for reaching the goal state (i.e., time to completion). There was no time limit on the task.

The task began with two one-move items and one two-move practice item that allowed children to become familiar with how to move the monkeys from one tree to another. Following the practice set, one item requiring a minimum of two moves was administered to begin the scored task. Though it was not anticipated that there would be major differences in performance on 1- and 2-move items, beginning with these items allowed for some degree of success by the 3-year-olds, which was felt to increase interest and motivation for "the game" and more complex items that were administered later (Bull et al., 2004). Two items were then administered for each of the 3-, 4-, 5-, 6-, and 7-move sets, respectively (see Appendix, Figure 1). Additionally, the number of GLGC moves required for each move increased across the problems, ranging from 0 to 3. There were 3 problems scored for the 0-, 1-, and 2-GLGC items, and 2 problems scored for the 3-GLGC items (see Appendix, Figure 2). Goal states between the two towers were made comparable, being typically constructed in the tower-ending hierarchy on the far right peg (while this reverses traditional ToL presentation, it makes the two Tower tasks more visually similar). Due to mathematical limitations in matching move-count/GLGC combinations, three ToL items had to deviate from the tower-ending goal state and were instead given a partial-tower-ending goal state, given that previous studies have found this to be the less ambiguous than the flat-ending goal states (Kaller et al., 2011). On each item, participants were allowed to make the minimum number of moves plus an additional six moves before the item discontinued with a 'sign' indicating 'Let's Try Again' (e.g., for a 5-move item, children were allowed 5 plus 6 additional moves, for a total of 11 allowable moves before discontinuation). The entire task was discontinued if the child failed to complete two consecutive items of the same minimum move length. Given that each trial was more difficult than its preceding, discontinue criteria of fails in a row seemed warranted and mimics other cognitive measures (e.g., Digit Span tasks).

Two different primary outcome measures were calculated in order to look for possible differences that could occur in Tower correlation based upon scoring methodology. First, an efficiency score was calculated by dividing the number of minimum moves across the entire task by the total number of moves that the participant made on all items. For items in which a participant failed to reach the goal state in the maximum number of allowable moves, a participant received a score of maximum moves plus one (e.g., on an item with a maximum number of moves of 9, the participant would be assigned a 10, which ensured that those who solved the task on the final allowable move would not be equated with those who failed to solve that item). Because traditional Tower tasks use multiple trials in an effort to allow participants to try to become as efficient as possible for that move sequence (i.e., to demonstrate efficiency), this score was devised as a way to try to measure the efficiency component of Tower tasks. The second scoring method utilized was similar to more traditional approaches to scoring Tower tasks, in that it is a simple total score which awards a single point for every item successfully completed (out of 10).

2.3b- Tower of London (ToL) – Modified NEPSY Tower

The ToL task followed the format of the standardized NEPSY Tower task (Korkman, Kirk, & Kimp, 1998), using similar colored squirrels in trees to introduce the task, but was administered in a computerized version versus a paper and physical ToL format. All administrative and other non-structural components of the task were made to be identical to the modified ToM described above, with two key differences. One, this task required the use of a split screen to display the required goal state, being that three of the ten items had to deviate from the typical Tower-ending position of the ToM. This was due to limited start and goal states that can be made for each move count/GLGC combination that were being matched between the two Tower tasks. The task was introduced using the story instructions from the NEPSY of three differently colored, yet identically sized, squirrels (see Figure 4).

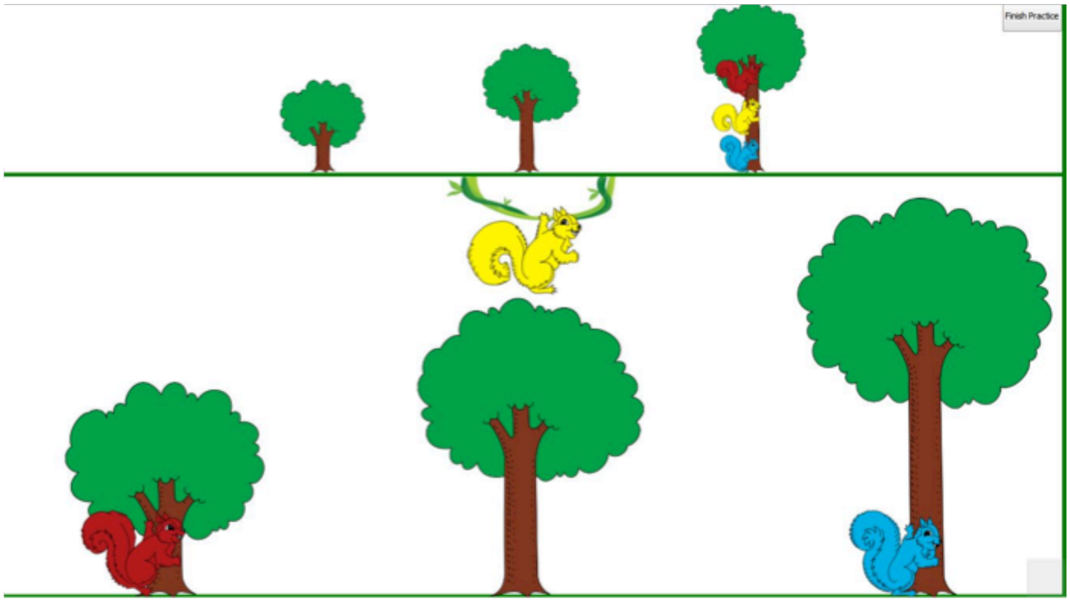


Figure 4- Tower of London Screenshot

Screen shot of the Tower of London (ToL) computerized measure created for this study, based upon the (ToL) from the NEPSY. In this version, a goal state screen was used to show children what the goal state would be, given that three of the ten items had to use alternative goal states. In this picture, the typical (i.e., 7 out of 10 items) goal state from the ToL is displayed at the top of the screen for the child, and he or she moves the squirrels in the bottom portion of the screen to try to match it.

This is in line with the traditional ToL, which utilizes three balls of identical size that are colored red, blue, and yellow. The background story from the same source will be used to provide comparable context to children, as does the ToH monkey story. The second major difference is that the tree sizes were adjusted to be in line with the traditionally graduated sizes of the ToL task. The trees were only able to hold one, two, and three squirrels, respectively. Children will be asked to indicate how many squirrels are allowed on each tree as a sign of understanding. At this point, all of the administrative, problem-structure, and scoring procedures are designed to be identical to the ToH (see above). In addition, the program collected total efficiency scores (i.e., total minimum moves across all items, divided by the total number of moves made by the participant)

2.3c- Go/No-Go task

The Dog-Koala go/no-go task is a touch-screen based computerized measure (Kerns & McInerney, 2007) inhibitory control task that requires the child to either respond (target) or to withhold a response (non-target) when a stimulus appears on the screen. The task begins with a baseline block in which the participant is told that pictures of a puppy will show up on the screen, and each time they see a picture of a dog they should touch it. Prepotent responses to the dog stimuli are established using this block. Following the baseline block, three more blocks are presented in which the participant are told to touch the screen whenever the established target stimuli for that block (e.g., for the three blocks, a koala, then dog, then koala, respectively) appears, and to not touch the screen whenever he or she sees the other animal (i.e., the non-target stimulus). Test blocks measure the number of omission errors (i.e., no response for target stimulus) or commission errors (i.e., response to non-target stimulus) and reaction time on all responses. The task takes approximately five minutes to administer.

2.3d- Boxes Task

The Boxes task (Kerns & McInerney, 2007) is screen touch computerized, self-ordered search task designed to measure working memory. Children are instructed that they will play a game to find a “Jack” in one of several boxes displayed on the computer screen. Once they find Jack in a box, he will hide again, but he will never hide in the same box twice, and they must continue to find Jack while playing the game. As such, participants must keep track of the boxes that they have already searched and in which they have already found ‘Jack’. During practice items, corrective feedback is given as needed. Upon completion of the practice items, children are then administered two items each of two, three, four, and five box sets. Though the rules are repeated at the beginning of each new set, corrective feedback is not provided once the test items begin. There are two types of possible errors for each item, within- and between-search errors. Within-item search errors are those in which the child checks the same empty box already looked in while trying to find Jack. Between-item search errors are those in which a child selects a box in which Jack has previously been found that game. Performance was measured by the average number of between-search errors across the four items of the 4- and 5-box conditions. The other conditions were excluded because they were hypothesized a priori to be either too simple (2- and 3-box conditions; i.e., visuospatial working memory might not be required for item success) or too difficult (6-box condition; i.e., strategized planning/problem solving may be abandoned when visuospatial WM demands are overloaded, which they were predicted to be a priori by this point in the task for preschool children).

2.3e- Flexible Item Selection Task (FIST)

The FIST (Jacques & Zelazo, 2001) is a measure of shifting/cognitive flexibility adapted from the Visual-Verbal Test (Feldman & Drascow, 1951) and designed for use with preschool children. Children were screened to ensure an understanding/knowledge of colors, sizes, and recognition of different objects that are used in the game. The investigator then played a practice game, where the child used one finger to select two (out of three available) favourite pictures that were all on one card that was placed on a table in front of the child. This was repeated three times to help teach the child to only use one finger to select the pictures on each card, and that they were only to touch two pictures per selection. Following the 'favorites' game, the investigator demonstrated the FIST task, with an explanation of the rules. Two practice items followed, wherein the children made the selections and received feedback from the investigator. After practice items, nine items were administered without any feedback provided. On each of the nine items, a paper was placed in front of the children that contained three different pictures. By design, the three pictures could be matched into pairs according to size (e.g., small, medium, large), object (e.g., boat, shoe, teapot), or colour (e.g., yellow, red, blue). In each three card set, there were two pairs of matching cards, meaning that each set of pictures had: 1) A target/shift picture that matched each of the other two pictures in some unique way (i.e., by size, object, or colour) from the other, and 2) two pictures that each matched the target/shift picture in some unique way, but that did not match each other in a unique way (e.g., if they only matched by both being blue, the target/shift card would have also been blue, meaning that these two pictures could not be selected as a pair that match in a 'unique' way).. After making the initial selection/match, children were then asked to pick two other pictures that were the same in a different way than the first pair. Children received one point if they were able to successfully pick two pairs of matching items on the task, which indicated that they had successfully 'shifted' their use/categorization of the target/shift picture.

3- Results

Data analyses were conducted in order to address three primary issues: (1) the relationship between the ToM and ToL and whether correlation in performance between these two tower tasks in preschoolers was significantly above the reported correlational range of previous Tower studies (.39-.6), as described by Welsh et al. (2000), (2) if the correlations differed between performance on the two Tower tasks for different levels of item complexity (i.e., simple vs. complex items), and (3) the relationship between tower performance and measures of inhibition, working memory, and shifting. Additional analyses addressed differences between two groups that emerged as a result of discontinuation patterns, which are described below.

Statistical analyses were completed using SPSS Version 21.0. The number of children who completed both tower tasks, and thus could be used in the comparison of scores on the tower tasks was 29 (14-boys, 15-girls, mean age=55 months). Calculation of Pearson correlation coefficients between the two total efficiency scores revealed a strong correlation between the two Tower tasks, r (29) = .668, p < .001. In order to compare this finding with Welsh et al.'s (1999) finding of r (35)= .39, Fisher's r-to-z transformation (as described by Cohen, 1983; originally proposed by Steiger, 1980) was utilized for comparing two correlation coefficients obtained on two different random samples. In order to accomplish this, the obtained r's were converted into z' equivalents (using Appendix Table B; Cohen & Cohen, 1983, pp. 520; r= .668➔z'=.811, and r=.39➔z'=.412) and used with their associated N's to calculate a score of z= 1.59. This value was then compared with the two-tailed α = .05 criterion (1.96, found in Appendix Table C of Cohen & Cohen, 1983, pp. 521), which gave a value of p = .055, which was insufficient to reject the null hypothesis that the obtained correlation of this study was statistically different than that of Welsh et al.'s (1999) obtained correlation. Further, using the same method for comparing correlation coefficients (Cohen & Cohen, 1983), the obtained correlation was not significantly above what Welsh et al. (2000) found with adults when they attempted to control for administrative, though not structural, differences between the ToH and ToL-R (z = -.04, *ns*). The two Tower tasks were further compared using total number of tower problems correctly completed, which also yielded a strong correlation of r (29) = .560, $p < .01$. Again when compared to Welsh et al.'s published r = .39 (1999); this result also failed to meet traditional significance levels (z = -0.93, *ns*).

Though overall results for the entire sample of children who completed the two Tower tasks revealed strongly correlated efficiency and total scores, there was substantial variability within the participants in interest and performance on the tasks, with many children failing to complete the entire task (e.g., stopped responding in the tower task before hitting the discontinue criteria or the end of the task). Of the 29 participants who attempted both Tower tasks, only 10 participants actually completed both tower tasks (e.g, continued both tasks until reaching the discontinuation rule of two consecutive failed items). The remaining 19 participants either quit one (N=9) or both (N=10) Tower tasks before they reaching the conclusion of the task. Given the substantial difference not finishing the task has on total performance comparison of equivalence in performance between the two tasks can be significantly affected. In previous studies examining the relationship between these two tower tasks, this type of participant attrition was not a factor (Welsh et al., 1999; 2000). To better understand the impact of approximately twothirds of the sample failing to correctly complete the tasks, the sample was divided into those participants who reached a ‘legitimate’ end to both Tower tasks (termed ‘non-quitters’ or ‘NQ group hereafter; i.e., ending the task only after failing two consecutive same-move items or completing the last item of the task;# of boys = 6, # of girls = 4, mean age= 57.8 months) and those participants who quit the task early (termed ‘quitters’ or ‘Q Group’ hereafter; i.e., those who refused to continue the task until failing two consecutive same-move items or reaching the end; # of boys= 8, # of girls= 11, mean age= 54.21 months). In order to determine whether gender was related to group placement, Fisher’s Exact Test was used (based on expected cell count) and revealed no significant relationship between gender and group (p= .450). Analyses conducted with this subsample of ‘non-quitters’ revealed even strongly correlations between the performance on the two tower tasks, and analyses revealed efficiency correlations of (r (10) = .768, $p < .01$) for efficiency and (r (10) = .811, $p < .01$) for total scores. Indeed, the total score correlation was significantly above Welsh et al.’s (1999) findings ($z = -1.72$, $p < .05$) even given the small sample size and thus reduced power to find a difference between the two correlation coefficients. However, the obtained total score correlation was not significantly higher than the highest adult correlations previously reported (r (16)= .66) which was obtained by Welsh et al. (2000; $z = -0.72$, *ns*).

To investigate the impact of item complexity on the correlation between Tower tasks performance, Pearson r’s were also calculated across simple (i.e., 0- and 1-GLGC move items) and complex (i.e., 2- and 3- GLGC move items) Tower problems in the “non-quitter” group. The analyses found that for simple problems, the two towers were not significantly correlated using either efficiency (r (10) = .599, *ns*) or total (r (10) =

.564, *ns*) scores. On complex Tower problems, however, there was a strong correlation for both efficiency (r (10) = .807, p <.01) and total (r (10) = .806, p < .01) scores (see Table 1). Given that neither the efficiency nor total scores on simple problems were correlated at a statistically significant level between the two Tower tasks, the method for calculating the significance of the differences between dependent r's, as outlined by Cohen & Cohen (1983), could not be used to compare the correlations at the two levels of complexity. However, given these results, these data suggests that for more complex problems, performance on the two Tower tasks is much more strongly associated. In addition to examining the impact of problem complexity on the equivalence of performance on the two Tower tasks, additional analyses were conducted to explore whether different types of goal-states (i.e., partial-tower vs. full-tower) impact performance, as suggested by previous research by Unterrainer and colleagues (2005). Paired-samples t-tests were conducted to compare the mean number of correct solutions across the three partial-tower goal-state items on the ToL, with the associated (i.e., matched for both move-count and counterintuitive moves) ToM full-tower goal-state items in the NQ group.

Table 1- Tower Task Correlations (NQ Group)

	ToM Eff	ToM Total	ToM Simp Eff	ToM Simp Tot	ToM Comp Eff.	ToM Comp Tot	ToL Eff	ToL Tot	ToL Simp Eff.	ToL Simp Tot	ToL Comp Eff
ToM Efficiency	1	-	-	-	-	-	-	-	-	-	-
ToM Total	.938**	1	-	-	-	-	-	-	-	-	-
ToM Simple Eff.	.935**	.875**	1	-	-	-	-	-	-	-	-
ToM Simple Tot.	.736*	.873**	.771**	1	-	-	-	-	-	-	-
ToM Complex Eff.	.971**	.913**	.824**	.657*	1	-	-	-	-	-	-
ToM Complex Tot.	.952**	.971**	.847**	.732*	.955**	1	-	-	-	-	-
ToL Efficiency	.768**	.752*	.782**	.613	.704*	.752*	1	-	-	-	-
ToL Total	.829**	.811**	.783**	.567	.808**	.857**	.914**	1	-	-	-
ToL Simple Eff.	.489	.530	.599	.581	.376	.457	.843**	.626	1	-	-
ToL Simple Tot.	.771**	.770**	.762*	.564	.728*	.801**	.919**	.928**	.792**	1	-
ToL Complex Eff.	.816**	.755*	.751*	.501	.807**	.811**	.878**	.932**	.486	.791**	1
ToL Complex Tot.	.806**	.777**	.761*	.527	.786**	.828**	.900**	.996**	.587	.902**	.946**

ToM-ToL Tower Task correlations for efficiency and total scores, as well as each score for both simple (i.e., 0- and 1-GLGC move) and complex (i.e., 2- and 3-GLGC move) items for the Non-Quit group (N=10). *Note*: * $p < .05$; ** $p < .01$

Table 2- Tower Task Correlations (Q Group)

	ToM Eff	ToM Total	ToM Simp Eff	ToM Simp Tot	ToM Comp Eff.	ToM Comp Tot	ToL Eff	ToL Tot	ToL Simp Eff.	ToL Simp Tot	ToL Comp Eff
ToM Efficiency	1	-	-	-	-	-	-	-	-	-	-
ToM Total	.924**	1	-	-	-	-	-	-	-	-	-
ToM Simple Eff.	.926**	.892**	1	-	-	-	-	-	-	-	-
ToM Simple Tot.	.758**	.893**	.852**	1	-	-	-	-	-	-	-
ToM Complex Eff.	964**	.851**	.795**	.606**	1	-	-	-	-	-	-
ToM Complex Tot.	.918**	.939**	.796**	.684**	.917**	1	-	-	-	-	-
ToL Efficiency	.641**	.450	.479*	.253	.696**	.537*	1	-	-	-	-
ToL Total	.602**	.435	.484*	.265	.623**	.503**	.929**	1	-	-	-
ToL Simple Eff.	.550*	.315	.432	.161	.595**	.388	.837**	.698**	1	-	-
ToL Simple Tot.	.483*	.277	.415*	.142	.486*	.341**	.776**	.871**	.795**	1	-
ToL Complex Eff.	.568*	.461*	.414	.283	.615**	.532*	.881**	.883**	.481*	.553*	1
ToL Complex Tot.	.598**	.453	.476*	.279	.620**	.522*	.933**	.988**	.639**	.791**	.941**

ToM-ToL Tower Task correlations for efficiency and total scores, as well as each score for both simple (i.e., 0- and 1-GLGC move) and complex (i.e., 2- and 3-GLGC move) items for the Q group (N=19).

Note: * $p < .05$; ** $p < .01$

The analysis revealed that, overall, there was not a significant difference in the total scores for the partial-tower ToL items (M= 1.6, SD= 1.174) and the full-tower ToM items (M=2.20, SD= .789); t(9)= -1.96, *ns*). However, a subsequent paired-samples t-test revealed that while children were able to complete approximately the same number of items correctly, there was a significant difference in their efficiency scores for the partial-tower ToL items (M= .531, SD= .105) versus full-tower ToM items (M= .664, SD= .193); t(9)= -2.798, $p < .05$. Thus it is possible that the more ambiguous partial-tower goal states resulted in subjects requiring more moves above the minimum amount to complete the items, and thus less efficiency on the TOL items.

Investigation of the 'quit early' group was undertaken to determine whether a similar pattern of performance would be seen, again conducting paired-samples t-tests were conducted for both total and efficiency scores on these alternative goal items. The analyses revealed that in this group, there was a significant difference in total scores for the partial-tower ToL items (M= 1.79, SD= 1.032) and the full-tower ToM items (M= 1.21, SD= 1.032); t (18)= -2.48, $p < .05$, as well as a significant difference in efficiency scores for partial-tower ToL items (M= .644, SD= .195) and the full-tower ToM items (M= .512, SD= .125); t (18)= -3.503, $p < .01$). These findings suggest that performance on these item partial-tower TOL items by subjects in this group also was negatively impacted.

Analyses were also completed to investigate the relationships between three tasks of executive function (i.e., inhibition, working memory, and shifting) and Tower task performance. Tables 3 and 4 provide means and standard deviations for the NQ and Q Groups, respectively. Tables 5 and 6 provide correlations between the five separate EF measures and the two Tower tasks for the two groups of participants, respectively. No significant relationships were observed between either of the Tower tasks and any executive function measure for children in the group who 'Quit' on Tower tasks (see Table 6). The relationships between EF measures and Tower task performance differed for the "Non-Quit" group showed no significant correlations between Tower performance and the Go/No-Go (inhibitory control), the Continuous Performance Test (sustained attention), or the Flexible Item Selection Task (shifting/cognitive flexibility), with the exception of the ToL Simple Total score which significantly correlated with performance on the Stroop (r= .676, $p < .05$: See Table 5). In contrast, the working memory measure, Jack's Boxes (between-error score), was strongly correlated to all but two of the Tower scores (ToL Complex Total, ToM Simple Total). On the two Tower total efficiency scores, the Boxes score correlated strongly with both the ToM (r= -760, $p < .05$) and the ToL (r= -907, $p < .01$). On the Tower total scores, Boxes also correlated strongly with the ToM (r= -.691, $p < .05$) and the ToL (r= -756, $p < .05$).

Table 3- NQ Group Means and Std. Deviations

Variable	Mean	Standard Deviation
Age (Months)	57.8 (M= 58.3, F= 57.0)	4.34 (M= 3.5, F= 5.8)
ToM Efficiency Score	.613	.119
ToM Total Score	6.5	2.42
ToM Simple Efficiency	.686	.172
ToM Simple Total	4.4	.843
ToM Complex Efficiency	.58	.102
ToM Complex Total	2.1	1.73
ToL Efficiency Score	.53	.058
ToL Total Score	4.9	2.18
ToL Simple Efficiency	.559	.103
ToL Simple Total	3.6	1.08
ToL Complex Efficiency	.517	.047
ToL Complex Total	1.3	1.25
Stroop Total Correct	17.55	1.86
GNG Total Score	5.0	4.57
CPT Total Score	6.44	4.28
Boxes BT-Errors	3.5	1.26
FIST Total Score	7.09	1.85

Means and Standard Deviations of the performance variables for the Tower, Inhibition, Sustained Attention, Working Memory, and Shifting tasks for the Non-Quit Group.

Note: ToM/ToL Efficiency Score (minimum moves divided by moves made, perfect score=1.0); ToM/ToL Total Score (10 possible); ToM/ToL Simple Efficiency (efficiency score out of the 0- & 1-GLGC move items, 5 total); ToM/ToL Complex Efficiency (efficiency rating out of the 2- & 3-GLGC move items, 5 total); ToM/ToL Simple/Complex Total (5 possible); Stroop Total Correct (20 possible); GNG/CPT Total Score (Omission/Commission errors combined); Boxes Between-Errors (Average between-error count per block, Blocks 4-5); FIST Total Score (9 possible)

Table 4- Q Group Means and Std. Deviations

Variable	Mean	Standard Deviation
Age (Months)	54.21 (M= 52.88, F= 55.18)	5.93 (M= 6.26, F= 5.78)
ToM Efficiency Score	.607	.135
ToM Total Score	5.74	2.68
ToM Simple Efficiency	.688	.202
ToM Simple Total	3.95	1.27
ToM Complex Efficiency	.571	.133
ToM Complex Total	1.79	1.65
ToL Efficiency Score	.536	.080
ToL Total Score	4.47	2.07
ToL Simple Efficiency	.568	.131
ToL Simple Total	3.47	1.07
ToL Complex Efficiency	.524	.078
ToL Complex Total	1.0	1.25
Stroop Total Correct	16.5	1.98
GNG Total Score	7.66	4.55
CPT Total Score	12.0	7.06
Boxes BT-Errors	3.17	1.43
FIST Total Score	7.13	1.43

Means and Standard Deviations of the performance variables for the Tower, Inhibition, Sustained Attention, Working Memory, and Shifting tasks for the Quit Group.

Note: ToM/ToL Efficiency Score (minimum moves divided by moves made, perfect score=1.0); ToM/ToL Total Score (10 possible); ToM/ToL Complex Efficiency (efficiency rating out of the 2- & 3-GLGC move items ToM/ToL Complex Efficiency (efficiency rating out of the 2- & 3-GLGC move items, 5 total); ToM/ToL Complex Total (5 possible); Stroop Total Correct (20 possible); GNG/CPT Total Score (Omission/Commission errors combined); Boxes Between-Errors (Average between-error count per block, Blocks 4-5); FIST Total Score (9 possible)

Table 5- Tower/ EF measure Correlation (NQ Group)

Variable	Stroop	GNG	CPT	Boxes	FIST
ToM Efficiency	.369	-.448	-.160	-.760*	-.345
ToM Total Score	.326	-.199	-.034	-.691*	-.386
ToM Simple Efficiency	-.308	-.392	-.070	-.771*	-.312
ToM Simple Total	.198	.248	.474	-.564	-.451
ToM Complex Efficiency	.378	-.455	-.221	-.693*	-.336
ToM Complex Total	.359	-.381	-.253	-.704*	-.319
ToL Efficiency	.442	-.627	-.189	-.907**	-.207
ToL Total Score	.574	-.606	-.327	-.756*	-.392
ToL Simple Efficiency	.466	-.462	.091	-.778*	-.112
ToL Simple Total	.676*	-.695	-.394	-.810**	-.154
ToL Complex Efficiency	.317	-.531	-.326	-.728*	-.329
ToL Complex Total	.421	-.456	-.231	-.597	-.551

Tower task correlations with the executive function measures for the Non-Quit Group

Note: * $p < .05$; ** $p < .01$

Table 6- Tower/ EF Measure Correlations (Q Group)

Variable	Stroop	GNG	CPT	Boxes	FIST
ToM Efficiency	-.048	.056	-.172	.026	-.208
ToM Total Score	-.116	.025	-.106	.016	-.031
ToM Simple Efficiency	-.127	.009	-.143	.026	-.178
ToM Simple Total	-.045	.037	.032	-.146	-.138
ToM Complex Efficiency	-.009	.088	-.169	.022	-.207
ToM Complex Total	-.153	.012	-.199	.121	.056
ToL Efficiency	.261	-.234	-.386	-.107	-.130
ToL Total Score	.251	-.096	-.383	-.023	-.192
ToL Simple Efficiency	.286	-.333	-.403	.099	-.276
ToL Simple Total	.287	-.116	-.370	.224	-.342
ToL Complex Efficiency	.161	-.089	-.282	-.281	.021
ToL Complex Total	.168	-.060	-.318	-.234	-.024

Tower task correlations with executive function measures for the Quit Group

Note: No significant correlations were found in this group

In order to better understand what differences might exist between the Quit and Non-Quit groups, and whether the differential pattern of performance may be attributable to other factors, post-hoc independent-sample t-tests were used to test for differences between the two groups in regards to age, Tower task performance, and executive function task scores. The two groups were not found to be significantly different in age, ToM/ToL performance (across all Tower scores and complexity levels), or on the primary measures of executive function (i.e., Stroop, Go/No-Go, Boxes, and Flexible Item Selection Task). This suggests that differences between the two groups are unlikely to be attributable to development, a higher proficiency on Tower tasks, or increased inhibition, working memory, or shifting/cognitive flexibility. However, on the Continuous Performance Test (CPT), a measure of sustained attention, the Q group (M= 11.43, SD= 6.08) was found to perform significantly lower than the NQ group (M= 6.78, SD= 4.17); $t(27) = 2.155$, $p < .05$. This suggests that, while there was not a significant difference in mean performance on the Tower tasks, the Q group may have been more susceptible to losing interest or focus during one or both of the Tower tasks and subsequently quitting the task before reaching a ‘legitimate’ discontinue. Given that there was not a significant difference in mean performance on Tower tasks between the two groups, but other differences were observed (e.g., correlation with the WM task), these data suggests that decreased sustained attention may have caused enough variability in effort, performance, and task completion so as to confound the results for the Q group, perhaps secondary to lower reliability of the tower tasks (based on fewer observations) or participants true ability to problem-solve (versus sustain attention to task).

4- Discussion

Though a number of previous studies have analyzed Tower task performance in young children (Bull et al., 2004), it was unclear whether eliminating administrative and structural differences between two Tower tasks would increase their correlation in a population in which problem solving abilities are still emerging. The purpose of the current study was to examine whether the relationship between performances on two Tower tasks is similar (or 'higher') in preschool children than has previously been reported in adults, after controlling for administrative and structural differences between the two tasks. Previous studies investigating the correlation between Tower of Hanoi (ToH) and Tower of London (ToL) have generally reported findings in the range of.27 to .66 (Humes et al., 1997; Welsh, et al., 1999; 2000; Zook et al., 2004). Past studies have had mixed results in Tower comparisons, with some studies finding no difference between results in Towers tasks when attempting to eliminate administrative and/or structural differences (Welsh et al., 2000) and others reporting significant increases in Tower correlations (Kaller et al., 2004; Unterrainer et al., 2005). Though a number of previous studies have analyzed Tower task performance in young children (Bull et al., 2004), it was unclear whether eliminating administrative and structural differences between the two tasks would increase their correlation in a population in which problem solving abilities are still emerging.

Though initial analyses using the entire sample (N=29) found a strong correlation between the two Tower tasks, the correlation coefficents from both efficiency and total scores in the entire sample were not statistically different than previously reported low correlation findings. However, it is important to note that there are key differences between Tower task administration for the preschoolers and that of previous adult Tower studies. Namely, in most adult studies, participants were required to complete all items for the Tower tasks (i.e., there was no Tower task discontinuation rule). In the current study, however, only one-third (10) of the participants were actually willing or able to complete both Tower tasks until they either reached the games end by discontinuation (i.e., two failed items of the same move-count in a row) or by completing all items. Given that discontinuing either Tower task could greatly alter the scores, based on when the child decided to no longer continue 'play the tower game), scores for children who did not complete the tasks may underestimate children's problem solving skills, and will also reduce the reliability of the tasks, limiting the correlations between them. As such, the sample was split into two different groups, those who completed all of the items ("Non-Quit"/NQ group) and those who quit on one or both of the Tower tasks ("Quit"/Q group). Further exploration of the data, examining only the Non-Quit group was therefore used for answering the primary questions of the study, as this group's data is not impacted by these concerns. Also, in order to better understand how differences in participant characteristics and task discontinuation can impact Tower task correlation, the Quit group was included in some analyses for exploratory value.

When analyzed separately, the NQ group did show large correlations for both efficiency and total scores. However, while these correlations were significantly above the lower-end of previously reported Tower task correlations, neither was significantly above the correlation range that was reported by Welsh et al. (1999). While the low number of children who completed both Towers contributes to the lack of statistical difference between the observed preschool Tower task correlation obtained in this study vs. previous research, without sufficient power it is unclear whether the strength of the correlation is truly different. Therefore, while this study provides some support for increased Tower task correlation in preschool children completing Tower tasks matched for administrative and structural factors, further investigation is still needed.

4.1- Impact of Task Complexity

As hypothesized, problem complexity does appear to impact the correlation between the two Tower tasks. Interestingly, simple problem measures failed to produce significant correlations, possibly due to a lack of variability in simple item scores that resulted from the relative simple nature of 0- and 1- GLGC problems (e.g. a ceiling affect with most children getting all items correct). In contrast, all of the complex problem measures were strongly correlated between the two Tower tasks, suggesting that on items with increased difficulty (2- and 3-GLGC problem items and higher move-count items) may be tapping actual problem-solving/planning abilities, rather than participants being able to solely rely upon perceptually-driven approaches for item completion (i.e., as more simplistic items allow; see Kaller, et al., 2004, 2011; Unterrainer, et al., 2008). However, given that these items also had a greater variance in performance within the sample, it is difficult to determine whether these items are better measuring problem-solving skills or just provide a better range of scores to support the noted correlations.

4.2- Impact of Goal State Ambiguity

While it was not the explicit intention of the study to assess difference in performance between items that have either Partial-Tower or Full-Tower goal states, there were three (out of ten total) ToL items that were Partial-Tower goal states. Previous studies have shown that the ambiguity that is associated with Partial- and Flat-ending

Tower items comes from a decreased ability to determine which disk/ball (or, in this case, monkey/squirrel) must be placed into the ‘final’ position first (Kaller et al., 2004, 2011; Klahr, 1985). A possible result of this ambiguity comes when a participant follows a suboptimal solution path, meaning that they might incorrectly place a disk/ball into its final position before they should, and have to backtrack in order to solve the problem. Depending upon the specific parameters of a given Tower study, this may result in failing that item (i.e., if the Tower task requires a participant to reach the goal state in the minimum number of moves), receiving a lower score (i.e., for those Tower tasks that utilize an efficiency score, allowing for certain number of additional moves beyond the minimum requires, as was the case for this study), or make no difference at all (e.g., if the suboptimal solution path was a small deviation and the participant is able to reach the goal state before they reach the maximum allowable moves).

Interestingly, the two groups in this study showed slightly different outcomes in response to the Partial-Tower goal state items. The NQ group showed no difference between Partial- and Full-Tower goal state items on their total scores, indicating that they were able to successfully complete both types of Tower problems about equally well regardless of goal stated. However, when efficiency scores were used to compare the two types of problems, the NQ group showed a significantly lower efficiency in completing these Partial-Tower goal state items. This suggests that these children were more likely to take a suboptimal solution path to the goal state, but that they were eventually able to reach the goal. In contrast, the Q group showed significant differences in both total scores and efficiency scores on Partial-Tower goal state items, indicating that they were more likely to be discontinued on these items and, even if they did reach the goal state, they required significantly more moves to do so. Taken together, these findings suggest that goal states in Tower task comparison studies should be matched for their level of ambiguity (i.e., Flat-, Partial-, and Full-tower ending goal states). Furthermore, these findings provide additional support for using different types of goal states to increase difficulty on Tower tasks (Kaller et al., 2004, 2011; Klahr, 1985).

4.3- Tower task correlation with other EF measures

For both the ToM and the ToL, the working memory measure (Jack's Boxes; total between-error score) was very highly correlated with overall Tower performance. A number of previous studies have failed to consistently find EF measures that are correlated with Tower performance, especially across both Tower tasks (Bull et al., 2004; Welsh et al., 1999). The specific EF measure that is selected is particularly important here, as the type of working memory (for example) that is tapped into should be as closely related to the type of working memory required for the particular Tower version being used. In this case, Jack's Boxes taps into visuospatial working memory using the same task interface for responding (i.e., touchscreen computer) as the ToM and ToL tasks utilized. This reduction in outside 'noise' (i.e., non-EF components to the task that can cause a reduction in the observed correlation) may have been a crucial aspect to capture the correlation between working memory and performance on these problem solving tasks. Alternatively, almost no other EF measures were significantly related to Tower performance (the only exception being a relationship between the Stroop and ToL Simple item performance) in spite of similar response interface on several of the measures. For the FIST, it is possible that a ceiling effect negated any potential relationship between the problem solving task and shifting abilities (e.g., 7 of 29 participants received a perfect score of 9, with 18 of 29 scoring 7 or above), and perhaps a more challenging task could have detected a correlation between the two EFs.

4.4- Individual differences between study groups

Given that there was such a significant difference between the Q and NQ groups, it was important to eliminate the possibility that these differences were simply attributable to age, or an increased ability to perform these tasks. However, this was not the case, as task performance across both Towers was not found to be significantly different between the two groups. Additionally, the two groups did not perform differently on the measures of inhibition, working memory, or shifting. The important difference between the groups was found on the CPT, a measure of sustained attention, where the Q group was found to perform significantly worse than the NQ group. This suggests that the decreased correlation between the two Tower tasks, and their lack of correlation with other EF measures (particularly the Boxes, given that this was highly related to performance in the other group), was likely attributable to their decreased ability to maintain their attention long enough on these tasks to perform consistently across the various measures. Their discontinuation from the Tower tasks were more random, meaning that it might have been unrepresentative of their ‘true’ problem solving abilities. This finding suggests that it might be individual differences in sustained attention that limit the correlations on similar Tower comparison studies in children.

4.5- Limitations & Future Directions

There were several limitations in this study that limited the findings. First of all, due to the fact that the sample split into two small groups (N=19 and N=10), there was limited statistical power and therefore some correlations that might actually be present may have not been big enough to be statistically significant given the small samples. Additionally, some statistical analyses were unable to be conducted, such as the formula for detecting a difference between dependent r's, which was invalid due to a small N. Secondly, while alternative goal states (i.e., Partial-Tower goal states on the ToL) had to be used to match the ToM on move-count and GLGC-moves, they introduced some variation between the two tasks that may have limited the size of the correlation. Participant's scores, particularly in the Quit Group, were negatively impacted by the presence of these three items. Additionally, as mentioned previously, the shifting measure was likely too easy, which caused there to be very little variability in total scores and, as a result, correlations with other measures.

While there are a number of advantages in using a touchscreen tablet with young children (e.g., appeal to children, ease of use, convenience), there are also some disadvantages. For example, some children responded impulsively while playing the games, such as when a child rapidly and repetitively taps on one of the stimuli (monkey or squirrel) in the Tower tasks. As this paradigm is currently programmed, the monkey/squirrel is moved back-and-forth from the tree to the vine repeatedly whenever a child engages in this repetitive tapping, which counts each back-and-forth sequence as one move. These repetitive taps were observed to occur for two reasons; 1) child confusion in trying to move the stimuli (though a 'stop' sign appears when an error movement is made, no such interference is made if the child moved the stimulus back-and-forth between the tree and the 'vine' that the ToM uses as a mechanism for moving stimuli, see Figure 3), and 2) frustration when the child is attempting to move a stimuli that is underneath another (e.g., attempting to move a monkey/squirrel that has another monkey/squirrel on top of it, prior to removing the obstacle, which can cause confusion for the child when the ‘wrong’ stimuli moves instead). The administrator watched for both of these types of errors, but it is impossible to keep them from occurring. Future studies using a more intuitive ‘touch-and-drag’ version of the Tower tasks (i.e., compared to this study’s move system that required an intermediate step, or use of the “vine”) should re-assess performance on the Tower tasks to determine whether ease of movement reduces error, and subsequently improves performance and increases the correlation of the two Tower tasks.

Bibliography

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Mikiewicz, O. (2002). Relationships between cognitive and behavioral measures of executive function in children with brain disease. *Child Neuropsychology*, *8*(4), 231-240. doi:10.1076/chin.8.4.231.13509

Anderson, P., Anderson, V. A., & Lajoie, G. (1996). The Tower of London test: Validation and standardization for pediatric populations. *The Clinical Neuropsychologist, 10(*1), 54-65. doi: 10.1080/13854049608406663

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychological Bulletin*, *121*(1), 65.

Bishop, D. V. M., Asmodt, G. A., McGurk, R., & Skuse, D. H. (2001). Individual differences in cognitive planning on Tower of Hanoi task: neuropsychological maturity or measurement error? *Journal of Child Psychology and Psychiatry. 42*(4), 551-556.

Borys, S. V., Spitz, H. H., & Dorans, B. A. (1982). Tower of Hanoi performance of retarded young adults and nonretarded children as a function of solution length and goal state. *Journal of Experimental Child Psychology*, *33*(1), 87-110. doi:10.1016/0022-0965(82)90008-X

Brocki, K. C., & Tillman, C. (2014). Mental set shifting in childhood: The role of working memory and inhibitory control. *Infant and Child Development*, doi:10.1002/icd.1871

Bull, R., Espy, K., & Senn, T. E. (2004). A comparison of performance on the Towers of London and Hanoi in young children. *Journal Of Child Psychology And Psychiatry*, *45*(4), 743-754. doi:10.1111/j.1469-7610.2004.00268.x

Bull, R., Espy, K., Wiebe, S. A., Sheffield, T. D., & Nelson, J. (2011). Using confirmatory factor analysis to understand executive control in preschool children: Sources of variation in emergent mathematic achievement. *Developmental Science*, *14*(4), 679-692. doi:10.1111/j.1467-7687.2010.01012.x

Byrnes, M. M., & Spitz, H. H. (1979). Developmental progression of performance on the Tower of Hanoi problem. *Bulletin of The Psychonomic Society*, *14*(5), 379-381.

Carder, H. P., Handley, S. J., & Perfect, T. J. (2004). Deconstructing the Tower of London: Alternative moves and conflict resolution as predictors of task performance. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *57A*(8), 1459-1483. doi:10.1080/02724980343000864

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. Developmental Neuropsychology, 29, 595–616.

Carlson, S.M. (2012). Executive Function Scale for Preschoolers. Test Manual, Institute of Child Development, University of Minnesota.

Chan, R. K., Shum, D., Toulopoulou, T., & Chen, E. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201-216. doi:10.1016/j.acn.2007.08.010

Chevalier, N., Sheffield, T. D., Nelson, J., Clark, C. C., Wiebe, S. A., & Espy, K. (2012). Underpinnings of the costs of flexibility in preschool children: The roles of inhibition and working memory. *Developmental Neuropsychology*, *37*(2), 99-118. doi:10.1080/87565641.2011.632458

Davidson, M. C., Amso, D., Anderson, L., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037-2078. doi:10.1016/j.neuropsychologia.2006.02.006

Diamond, A., Carlson, S. M., & Beck, D. M. (2005). Preschool Children's Performance in Task Switching on the Dimensional Change Card Sort Task: Separating the Dimensions Aids the Ability to Switch. *Developmental Neuropsychology*, *28*(2), 689-729. doi:10.1207/s15326942dn2802_7

Dick, A., & Overton, W. F. (2010). Executive function: Description and explanation. In B. W. Sokol, U. Müller, J. M. Carpendale, A. R. Young, G. Iarocci (Eds.) , *Self and social regulation: Social interaction and the development of social understanding and executive functions* (pp. 7-34). New York, NY, US: Oxford University Press.

Feldman, M. J., & Drasgow, J. (1951). A visual-verbal test for schizophrenia. *Psychiatric Quarterly Supplement*, *25*55-64.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal Of Experimental Psychology: General*, *137*(2), 201-225. doi:10.1037/0096-3445.137.2.201

Garon, N. M., Longard, J., Bryson, S. E., & Moore, C. (2012). Making decisions about now and later: Development of future-oriented self-control. *Cognitive Development*, *27*(3), 314-322. doi:10.1016/j.cogdev.2012.05.003

Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in 'planning' functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia*, *33*(5), 623-642. doi:10.1016/0028-3932(95)90866

Hughes, C., & Ensor, R. (2007). Positive and protective: Effects of early theory of mind on problem behaviors in at-risk preschoolers.*Journal Of Child Psychology And Psychiatry*, *48*(10), 1025-1032.

Hughes, C., Ensor, R., & Marks, A. (2010). Individual differences in false belief understanding are stable from 3 to 6 years of age and predict children's mental state talk with school friends. *Journal Of Experimental Child Psychology*, *108*(1), 96-112. doi:10.1016/j.jecp.2010.07.012

Humes, G. E., Welsh, M. C., Retzlaff, P., & Cookson, N. (1997). Towers of Hanoi and London: Reliability of two executive function tasks. *Assessment*, *4*(3), 249-257.

Jacques, S., & Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology*, *20*(3), 573-591. doi:10.1207/875656401753549807

Jostmann, N. B., & Gieselmann, A. (2014). When you have to climb downhill to reach the top: The effect of action versus state orientation on solving a goal-subgoal conflict in the Tower of Hanoi task. *Experimental Psychology*, *61*(5), 394-401. doi:10.1027/1618-3169/a000260

Kaller, C. P., Unterrainer, J. M., Rahm, B., & Halsband, U. (2004). The impact of problem structure on planning: Insights from the Tower of London task. *Cognitive Brain Research*, *20*(3), 462-472. doi:10.1016/j.cogbrainres.2004.04.002

Kaller, C. P., Rahm, B., Spreer, J., Mader, I., & Unterrainer, J. M. (2008). Thinking around the corner: The development of planning abilities. *Brain And Cognition*, *67*(3), 360-370. doi:10.1016/j.bandc.2008.02.003

Kaller, C. P., Rahm, B., Köstering, L., & Unterrainer, J. M. (2011). Reviewing the impact of problem structure on planning: A software tool for analyzing Tower tasks. *Behavioural Brain Research*, *216*(1), 1-8. doi:10.1016/j.bbr.2010.07.029

Kerns, K. A., & McInerney, R. (2007). Preschool Tasks [Computer software]. Victoria, BC, Canada: University of Victoria.

Klahr, D. (1985). Solving problems with ambiguous subgoal ordering: Preschoolers' performance. *Child Development*, *56*(4), 940-952. doi:10.2307/1130106

Klahr, D. (1994). Discovering the present by predicting the future. *The development of future-oriented processes* (pp. 177-218). Chicago, IL, US: University of Chicago Press.

Klahr, D., & Robinson, M. (1981). Formal assessment of problem-solving and planning processes in preschool children. *Cognitive Psychology*, *13*(1), 113-148. doi:10.1016/0010-0285(81)90006-2

Korkman, M., Kirk, U., & Kemp, S. (1998). NEPSY—A developmental neuropsychological assessment. San Antonio, TX: The Psychological Corporation.

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal Of Developmental Psychology*, *21*(1), 59-80. doi:10.1348/026151003321164627

Lezak, M. (1995). *Neuropsychological assessment (3rd ed.)*. New York, NY, US: Oxford University Press.

MacAllister, W. S., Bender, H., Whitman, L., Welsh, A., Keller, S., Granader, Y., & Sherman, E. S. (2012). Assessment of executive functioning in childhood epilepsy: The Tower of London and BRIEF. *Child Neuropsychology*, *18*(4), 404-415.

Martin, J., & Failows, L. (2010). Executive function: Theoretical concerns. In B. W. Sokol, U. Müller, J. M. Carpendale, A. R. Young, G. Iarocci (Eds.). *Self and social regulation: Social interaction and the development of social understanding and executive functions* (pp. 35-55). New York, NY, US: Oxford University Press.

Mataix-Cols, D., & Bartrés-Faz, D. (2002). Is the use of the wooden and computerized versions of the Tower of Hanoi Puzzle equivalent?. *Applied Neuropsychology*, *9*(2), 117-120. doi:10.1207/S15324826AN0902_8

McCormick, T. & Atance, C.M. (2011). Planning in young children: A review and synthesis. *Developmental Review. 31*, 1-31.

Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development*, *13*(3), 395-423.

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

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions In Psychological Science*, *21*(1), 8-14. doi:10.1177/0963721411429458

Müller, U., Kerns, K. A., & Konkin, K. (2012). Test–retest reliability and practice effects of executive function tasks in preschool children. *The Clinical Neuropsychologist*, *26*(2), 271-287.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Oxford, England:Prentice-Hall.

Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent?. *Ergonomics*, *51*(9), 1352-1375. doi:10.1080/00140130802170387

Piaget, J. (1976). The grasp of consciousness. Cambridge, MA: Harvard University Press

Salnaitis, C. L., Baker, C. A., Holland, J., & Welsh, M. (2011). Differentiating Tower of Hanoi performance: Interactive effects of psychopathic tendencies, impulsive response styles, and modality. *Applied Neuropsychology*, *18*(1), 37-46. doi:10.1080/09084282.2010.523381

Schiff, R., & Vakil, E. (2015). Age differences in cognitive skill learning, retention and transfer: The case of the Tower of Hanoi Puzzle. *Learning And Individual Differences*, *39*164-171. doi:10.1016/j.lindif.2015.03.010

Schnirman, G. M., Welsh, M. C., & Retzlaff, P. D. (1998). Development of the Tower of London—Revised. *Assessment*, *5*(4), 355-360. doi:10.1177/107319119800500404

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

Sikora, D. M., Haley, P., Edwards, J., & Butler, R. W. (2002). Tower of London test performance in children with poor arithmetic skills. *Developmental Neuropsychology*, *21*(3), 243-254. doi:10.1207/S15326942DN2103_2

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, *7*(2), 268-288.

Steiger, J. H. (1980). Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research*, 15(3), 335-352. doi:10.1207/s15327906mbr1503_7

Unterrainer, J. M., Rahm, B. B., Kaller, C. P., Leonhart, R. R., Quiske, K. K., Hoppe-Seyler, K. K., & ... Halsband, U. U. (2004). Planning abilities and the Tower of London: Is this task measuring a discrete cognitive function?. *Journal Of Clinical And Experimental Neuropsychology*, *26*(6), 846-856. doi:10.1080/13803390490509574

Unterrainer, J. M., Rahm, B. B., Halsband, U. U., & Kaller, C. P. (2005). What is in a name: Comparing the Tower of London with one of its variants. *Cognitive Brain Research*, *23*(2-3), 418-428. doi:10.1016/j.cogbrainres.2004.11.013

Unterrainer, J. M., Rauss, K. S., Kaller, C. P., Leonhart, R. R., & Rahm, B. B. (2008). Perceived conflicts and errors in complex problem solving. *Journal Of Clinical And Experimental Neuropsychology*, *30*(7),

Ward, G., & Allport, A. (1997). Planning and problem-solving using five-disc Tower of London task. *The Quarterly Journal Of Experimental Psychology A: Human Experimental Psychology*, *50A*(1), 49-78.

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal Of Educational Psychology*, *102*(1), 43-53. doi:10.1037/a0016738

Welsh, M. C., Revilla, V., Strongin, D., & Kepler, M. (2000). Towers of Hanoi and London: Is the nonshared variance due to differences in task administration?. *Perceptual And Motor Skills*, *90*(2), 562-572. doi:10.2466/PMS.90.2.562-572

Welsh, M. C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: Contribution of working memory and inhibition to performance. *Brain And Cognition*, *41*(2), 231-242. doi:10.1006/brcg.1999.1123

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure.. *Developmental Psychology*, *44*, 575–587.

Wiebe, S. A., Sheffield, T., Nelson, J., Clark, C. C., Chevalier, N., & Espy, K. (2011). The structure of executive function in 3-year-olds. *Journal Of Experimental Child Psychology*, *108*(3), 436-452. doi:10.1016/j.jecp.2010.08.008

Williams, D. J., & Noyes, J. M. (2007). Effect of experience and mode of presentation on problem solving. *Computers In Human Behavior*,*23*(1), 258-274. doi:10.1016/j.chb.2004.10.011

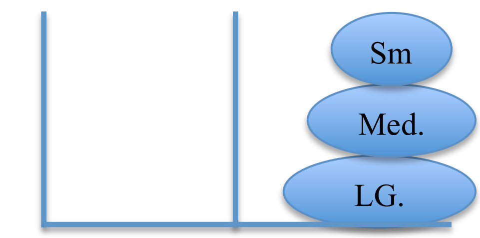
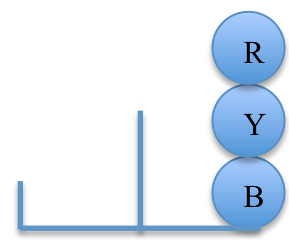
Zelazo, P., & Müller, U. (2011). Executive function in typical and atypical development. In U. Goswami (Ed.) , *The Wiley-Blackwell handbook of childhood cognitive development (2nd ed.)* (pp. 574-603). Wiley-Blackwell

Zinke, K., Fries, E., Altgassen, M., Kirschbaum, C., Dettenborn, L., & Kliegel, M. (2010). Visuospatial short-term memory explains deficits in Tower task planning in high-functioning children with autism spectrum disorder. *Child Neuropsychology*, *16*(3), 229-241. doi:10.1080/09297040903559648

Zook, N. A., Davalos, D. B., DeLosh, E. L., & Davis, H. P. (2004). Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. *Brain And Cognition*, *56*(3), 286-292. doi:10.1016/j.bandc.2004.07.003

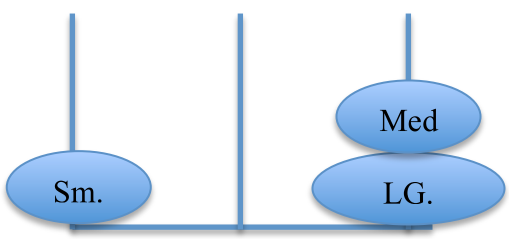
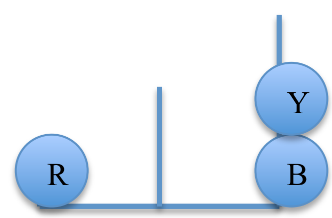
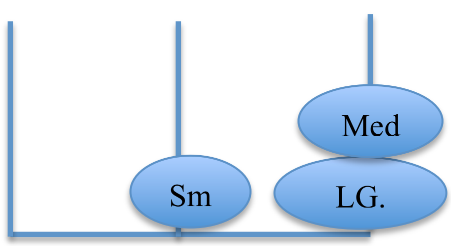
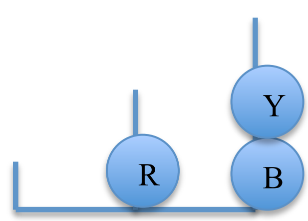
Appendix

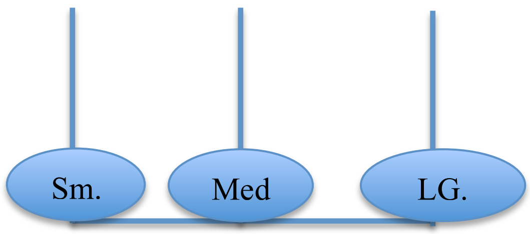
Appendix A- Tower Task Configuration

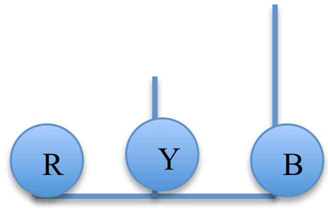
Primary Tower Goal States	
Tower of Monkeys (ToM)	Tower of London-Revised (ToL-R)
	

Sm.=Small, Med.=Medium, LG.=Large R=Red, Y=Yellow, B=Blue

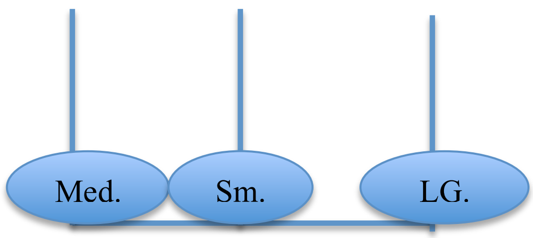
*Alternative goal-states outlined below.

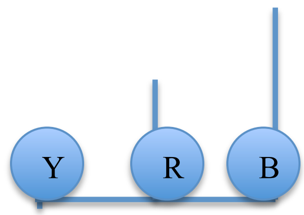
Tower Task Start States	
1- move, 0 GLGC	
	
1- move, 0 GLGC	
	
2- move, 0 GLGC	



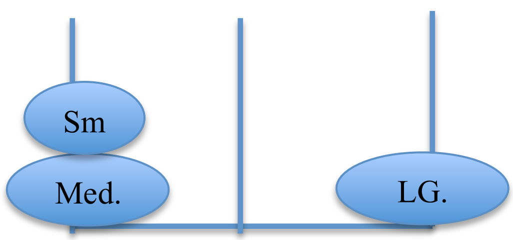


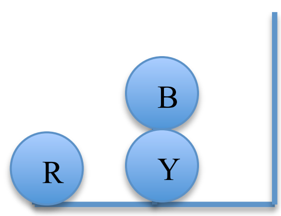
2- move, 0 GLGC



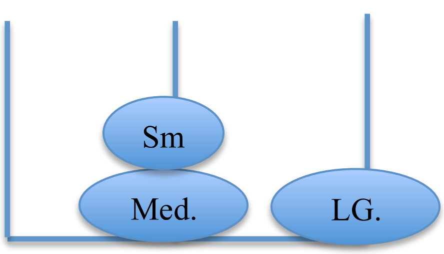


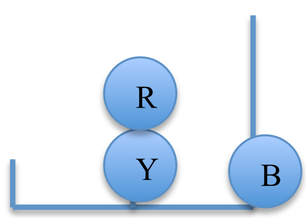
3-move, 0 GLGC





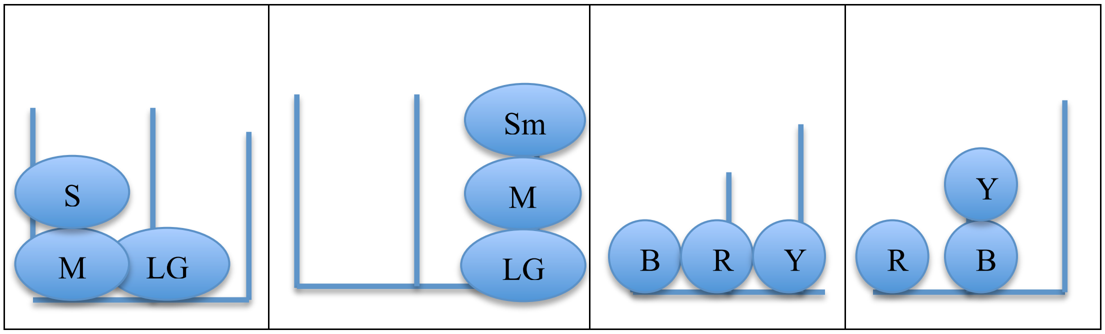
3-move, 1 GLGC



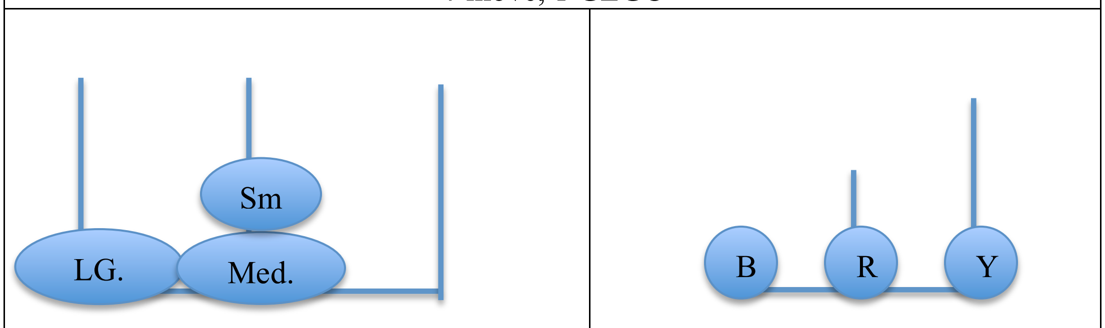


4-move, 0 GLGC

ToM Start	ToM Alt. Goal	ToL-R Start	ToL-R Alt. Goal

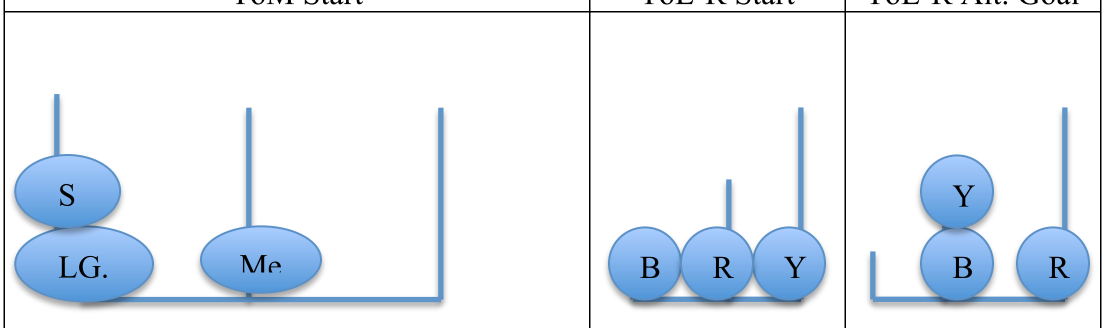


4-move, 1 GLGC

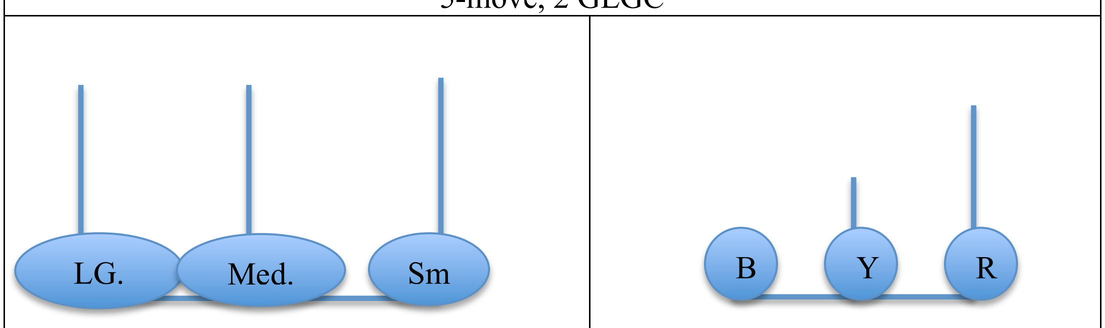


5-move, 1 GLGC

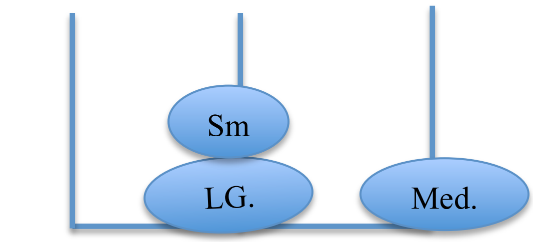
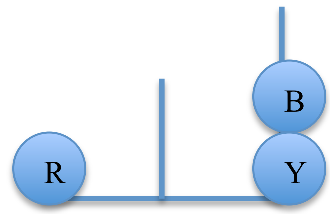
ToM Start	ToL-R Start	ToL-R Alt. Goal

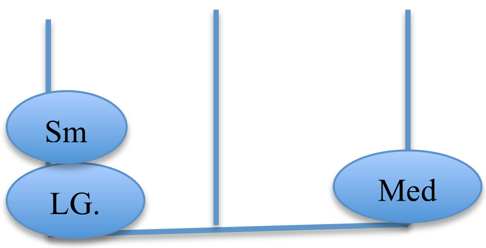
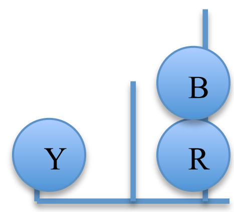
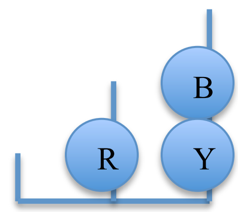
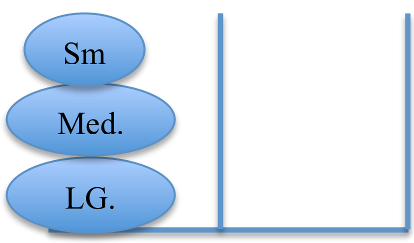
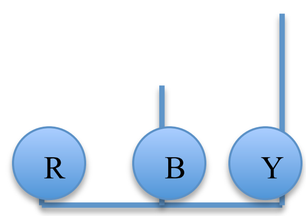
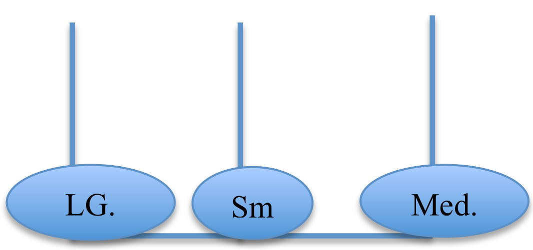
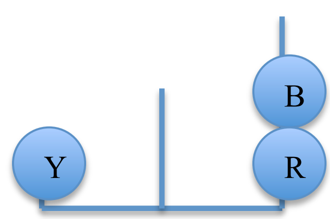


5-move, 2 GLGC



6-move, 2 GLGC

<table>
<tr><td></td><td></td></tr>
</table>

<table>
<tr><td colspan="3">6-move, 2 GLGC</td></tr>
<tr><td>ToM Start</td><td>ToL Start</td><td>ToL Alt. Goal</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td colspan="3">7-move, 3 GLGC</td></tr>
<tr><td></td><td colspan="2"></td></tr>
<tr><td colspan="3">7-move, 3 GLGC</td></tr>
<tr><td></td><td colspan="2"></td></tr>
</table>

Appendix B- Tower Task Organization by Problem Difficulty

Tower Task Organization by Problem Difficulty		
Simple vs. Complex	Move Count-GLGC#	Breakdown by GLGC #
Practice Items	1-0	0 Global/Local Goal Conflicts
	1-0	
	2-0	
Simple Items	2-0	
	3-0	
	4-0	
	3-1	1 Global/Local Goal Conflict
	4-1	
	5-1	
Complex Items	5-2	2 Global/Local Goal Conflicts
	6-2	
	6-2	
	7-3	3 Global/Local Goal Conflicts
	7-3	

Note: Tower problems in the computerized game were ordered sequentially by move count, and did not necessarily follow this exact order.

Appendix C: Telephone Screening Questions for Parents

1) Is your child within the age range of 3 to 5 years?
2) Because of the specific research questions that we are investigating, we need to ask a few questions about your child's developmental history. Is this all right with you?
3) Have you ever had any concerns about your child's development? For example, has your child had any services through the Infant Development Program, or had any difficulty with motor problems, language or speech, or vision?
   Note: Any child who received services/treatment for developmental concerns/delays was excluded from this study.
4) Does your child have any diagnosed medical, genetic or neurologic conditions?
   Note: Any child with a diagnoses such as Autism Spectrum Disorder, Noonan's Syndrome, and Selective Mutism were excluded from this study.
5) Has your child ever had a significant blow to the head (a fall or in an accident) that resulted in concern loss of consciousness, concern about concussion, or warranted medical investigation (such as an ER visit or visit with a doctor?
   Note: Any child who had experience the above was excluded from this study.