Out of the office and into the field: Exploring neuropsychological correlates between search behavior and a traditional desktop task in children and adolescents with ADHD

Marcos Francisco Rosetti,1,2 Rosa Elena Ulloa,3 Lino Palacios-Cruz,3 Robyn Hudson,1 Francisco de la Peña2

ABSTRACT

Introduction. Cognitive assessment of patients with attention deficit hyperactivity disorder (ADHD) can help clinicians provide individually tailored treatment and advice, and researchers to identify potential associations between psychopathology and specific cognitive deficits. Assessment instruments, however, have received some criticism regarding their ecological validity, that is, the capacity to extrapolate from the performance on such tasks to aspects of everyday functioning. In order to meet this challenge we developed the Ball Search Field Task (BSFT) that takes place outdoors and uses large, open areas. In the BSFT, the goal is to search for target objects hidden under opaque containers, with experimenters assessing the efficiency of participants’ strategies to collect a maximum of these. Objective. Here we explore how the measures produced by one of the latest versions of this task (the patchy BSFT) match up with a traditional desktop task often used in clinical environments, the Tower of London (ToLo). Method. We applied the BSFT and ToLo to children and adolescents with ADHD and compared the metrics using Spearman correlations. Results. We found significant, moderate correlations between instruments, as exemplified by that of balls collected per cones lifted (BSFT) and number of moves (ToLo) (r = -.44). Discussion and conclusion. Matching correlates between the BSFT and ToLo suggest these tasks may be tapping into similar cognitive processes. The addition of assessment tools with ecological validity may help provide a more comprehensive evaluation and a better understanding of the day-to-day impact of cognitive afflictions underlying psychiatric disorders such as ADHD.

Keywords: Neuropsychological tests, attention deficit hyperactivity disorder, Tower of London test, ecological validity, Ball Search Field Task.
INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental condition (American Psychiatric Association, 2013) with a large worldwide prevalence of over 5% (Polanczyk, Wilcutt, Salum, Kieling, & Rohde, 2014). The impact of ADHD on daily life can lead to reduced performance in social and academic settings (Barry, Lyman, & Klinger, 2002) and, later in life, at the workplace (Kessler, Lane, Stang, & Van Brunt, 2009).

Following an ADHD diagnosis, patients are often profiled using standardized neuropsychological tools. These procedures help clinicians gain deeper knowledge into cognitive impairments that may accompany the disorder to judge severity or provide treatment and advice tailored to each case (Brooks, Ploetz, & Kirkwood, 2016). In research, systematically testing clinical populations with such procedures has provided insight into some of the more common cognitive dysfunctions associated with ADHD and, in doing so, has aided efforts to identify potential neural circuits that may be affected (Castles, Kohnen, Nichols, & Brock, 2014). So far, most findings suggest a strong association between ADHD and deficits in executive function (Doyle, 2006; Nigg, Blaskey, Huang-Pollock, & Rappley, 2002; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005). Executive function (EF) is a broad term that encompasses a set of cognitive processes associated with the capacity to self-regulate behavior (Jurado & Rosselli, 2007), which include attention, inhibition, planning, and working memory. These cognitive functions are often interlinked, for instance, an attention deficit may influence working memory by interfering with obtaining and retrieving information.

A test often used to evaluate cognitive impairment in ADHD is the Tower of London (ToLo) (Shallice, 1982), which evaluates spatial problem solving abilities by having the participant rearrange a set of balls on wooden pegs to match a given pattern. In the ToLo, children with ADHD have been found to underperform, needing more moves to reach a solution, taking longer to complete some of the trials, and often handling the material in ways that are not allowed (Culbertson & Zillmer, 1998). A poor performance in the ToLo is interpreted by Culbertson and Zillmer (1998) as “a failure to plan and problem solve in a timely and rule-governed manner” (p. 221). A review by Seidman et al. (2004) neatly summarizes the results of neuropsychological studies in children, adolescents, and adults with ADHD. While many tasks have found a differential performance in ADHD subjects, not all have done so (Seidman et al., 2004), and differences are not always consistent across the lifespan, i.e. may be present in preschoolers but not in older children or adolescents (Culbertson & Zillmer, 1998), and even when these are present, effect sizes are often modest (Frazier, Demaree, & Youngstrom, 2004).

These difficulties in evaluating cognitive performance associated with ADHD point to the need for a paradigm shift in terms of evaluation procedures.

Heterogeneity in clinical manifestations of ADHD (e.g., comorbidity, severity, presence of cognitive disabilities or socioeconomic adversity) has made it challenging to draw clear links between neuropsychological scores and processes underlying the disorder (Culbertson & Zillmer, 1998). However, the nature of the assessment tools may also play a role. Criticisms of some of the findings showing ADHD subjects to underperform in neuropsychological tasks may relate to difficulties in separating issues of attention control from those of motivation (Burgess et al., 2006). One reason for this may be that most of the assessment tools are applied in a clinical setting and involve somewhat abstract desktop activities inside an office space. In a complimentary attempt to evaluate the behavioral performance of the children and adolescents with ADHD, we proposed to move out of the office and into the field by developing the Ball Search Field Task (BSFT) (Rosetti et al., 2016; Rosetti et al., 2018).

The BSFT attempts to address some of the issues often criticized in traditional tests by centering its design on the concept of ecological validity. By ecological validity we understand “the functional and predictive relationship between the patient’s performance on a set of neuropsychological tests and the patient’s behavior in a variety of real-world settings” (Sbordone, 1996, p. 16). The task is designed to simulate commonly faced situations such as looking around the house for a lost set of keys or navigating supermarket aisles in the attempt to locate a favorite food item. In such situations, people face several challenges; for example, avoid returning to previously visited locations (planning and memory), maintaining focus on the current activity until successful (attention and focus), and accessing recently acquired information that may point to the location of the desired item (working memory). In this sense, the BSFT evaluates cognitive function and behavior in a more naturalistic context as participants try to solve a biologically relevant “foraging” problem.

The BSFT involves an arrangement of opaque covers distributed over a large area, underneath of which a target object might be located. These elements allow the use of inexpensive, highly versatile setups. So far, we have tested three different versions of the BSFT by changing the spatial distribution of the targets and the size of the experimental arena (Rosetti et al., 2016; Rosetti, Valdez, & Hudson, 2017; Rosetti et al., 2018). Each version addresses different questions and potentially involves different cognitive processes. The first version consisted of a large arena on which the covers (brightly colored plastic cones) were arranged as a simple grid (Rosetti et al., 2016). Using this version, we showed that children with ADHD could collect the same number of target items as controls, but did so less efficiently. We then increased the difficulty of the task by changing the arrangement of the cones so that more planning was...
needed to reach an efficient solution (Rosetti et al., 2018), which proved useful to test adolescents. More recently, we modified the test by reducing the size of the search area and arranging the cones in a patchy distribution. In this last setup, participants need to detect the underlying distribution rule for the targets to improve their searching efficiency.

The present study explores the neuropsychological correlates of searching (planning and problem solving) in children and adolescents with ADHD by comparing measures of performance between the last version of the BSFT and the ToLo.

**METHOD**

**Participants**

We performed a convenience sampling of outpatients from the Child Psychiatric Hospital Juan N. Navarro in Mexico City from August 2017 to June 2018. Inclusion criteria were 1. a concurrent principal diagnosis of ADHD, 2. free of psychiatric medication for at least the previous six months, 3. showing no visible motor or sensory impairment, and 4. having an intellectual coefficient (IQ) of at least 80 points on the Wechsler Intelligence Scale for Children (Wechsler et al., 2003).

**Tools and measurements**

**BSFT**

This version consisted in five patches of 30 cones each (150 total), arranged over a flat concrete area of 6 x 9 m (Figure 1). Each patch contained six golf balls individually hidden as a cluster under six neighboring cones. The aim of the task was to collect into a cloth bag as many balls as quickly as possible (a maximum of eight minutes was allowed but not told to the participants). Instructions were a) to leave the cones upright after inspection and b) not to lift two cones simultaneously. Each participant’s sequence of collections was recorded using a small video camera (Hero4, GoPro, California, USA) mounted on a helmet worn by the participant, and later analyzed using the software BORIS (Friard & Gamba, 2016). Performance measures could be interpreted as proxies for planning, problem solving, sustained attention, and working memory. These measures are described in detail in Table 1.

To show efficient performance on the BSFT, the number of patches visited should be five, which would entail not repeating nor neglecting to visit any patches. Lifting many cones per patch suggests the participant was more focused on depleting the patch rather than extracting and applying the underlying rule. The number of balls collected is alone not necessarily indicative of a better performance. Rather, it is preferable to use measures such as collection rate (number of balls located per number of cones lifted) and the slope describing how collection rate changes as the participant explores more patches. A positive slope suggests the participant became more efficient as he or she visited successive patches. A smaller number of blank cones lifted after the last collection would suggest that the participant had gained information regarding the number of balls left in the patch. Similarly, a large probability of choosing a close cone after a collection suggests the participant was able to extract information regarding the distribution of balls and thus behaved as if expecting clumped resources.

**The Tower of London (ToLo)**

This task, designed by Shallice (1982) more than 35 years ago, consists of two sets of boards with three pegs each and three differently colored balls with holes so that they can be mounted on the pegs. The aim is for the participant to try and match the arrangement of balls as shown on the experimenter’s board using as few moves as possible. The task
consists of two sample trials followed by 10 scored trials. Performance measures are described in detail in Table 1. Overall performance is calculated by summing these variables across all trials and is interpreted as a proxy for planning and problem solving (Shallice, 1982).

A longer latency to the first move is interpreted as the participant taking time in order to plan the sequence of moves to execute. A larger number of correct sequences is considered a better performance than a smaller number of such moves. The presence of rule and time violations suggests poor performance due to the participant failing to pay attention to the task rules or taking too long to solve the task. Finally, shorter execution times often correlate with participant’s achieving a smaller number of moves (Shallice, 1982).

**Procedure**

The parents or guardians of the participating children were interviewed by an experienced clinical psychiatrist using a Spanish language version of the MINI-KID interview (Sheehan et al., 2010) to confirm diagnosis. Afterwards, participants were evaluated using the WISC-IV, and if the IQ inclusion criterion was met, continued to be tested on the two tasks. The order of presentation of the tasks was balanced, with a 20-minute pause between them.

**Statistical analyses**

To evaluate correlation values between the BSFT and the ToLo we calculated Spearman’s $r_s$. Using the `pcor.test` function from the `ppcor` package [24], we performed partial correlations to control for the effect of age for each comparison. We plotted all comparisons to visually corroborate correlation values. Significance for all tests was set at $p < .05$. All statistical tests were done using R (R Core Team, 2018).

**Ethical considerations**

All procedures were evaluated and approved by the internal review boards of the Institutions involved in the project. Written consent of parents or guardians and verbal consent of children were obtained before any evaluation was performed.

**RESULTS**

We recruited a total of 53 participants. From this sample, we discarded five BSFT participants: a) three failed to understand the instructions for the BSFT and stopped searching after finding only one ball in each patch, and b) two who ac-

<table>
<thead>
<tr>
<th>Performance measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of visited patches</td>
<td>The total number of patches visited. Participants could re-visit patches as often as they wished</td>
</tr>
<tr>
<td>Total number of lifted cones</td>
<td>The total number of cones lifted during the whole task from all patches visited</td>
</tr>
<tr>
<td>Total number of balls found</td>
<td>The total number of balls collected during the whole task</td>
</tr>
<tr>
<td>Overall collection rate</td>
<td>The total number of balls collected divided by the number of cones lifted</td>
</tr>
<tr>
<td>Slope of collection rate</td>
<td>The slope describing the linear relation between collection rate and the order in which patches were visited</td>
</tr>
<tr>
<td>Mean number of cones lifted after the last collection</td>
<td>The number of cones lifted after the last collection in a patch was made. A lower number of cones could indicate that the participant could assess the total number of balls per patch and thus choose to leave in a timely manner</td>
</tr>
<tr>
<td>Probability of choosing a close cone after a collection</td>
<td>The probability (percentage of choices of this type out of all choices) to select one of the nearest neighbor cones after a collection was made</td>
</tr>
<tr>
<td>Probability of choosing a close cone after an empty cone</td>
<td>The probability (percentage of choices of this type out of all choices) to select one of the nearest neighbor cones after lifting an empty cone</td>
</tr>
<tr>
<td>Latency to the first move</td>
<td>Time elapsed from the beginning of the trial until the participant makes the first move</td>
</tr>
<tr>
<td>Correct number of sequences</td>
<td>The number of trials in which the participant solved the task using the minimum number of moves</td>
</tr>
<tr>
<td>Number of moves</td>
<td>The number of times a ball was moved from one peg to another</td>
</tr>
<tr>
<td>Rule violations</td>
<td>The number of times the participant held more than one ball simultaneously or attempted to place more balls on an already full peg</td>
</tr>
<tr>
<td>Time violations</td>
<td>The number of trials in which the participant did not finish in the allotted time</td>
</tr>
<tr>
<td>Execution time</td>
<td>Time taken from the first move to the end of the trial</td>
</tr>
<tr>
<td>Total time</td>
<td>Sum of the latency to the first move and the execution time per trial</td>
</tr>
</tbody>
</table>
Neuropsychological correlates of search behavior in ADHD

Table 2
Spearman correlation values ($r_s$) expressing the level of association between the patchy version of the BSFT with ToLo ($n = 48$)

<table>
<thead>
<tr>
<th>BSFT</th>
<th>Number of patches</th>
<th>Total cones</th>
<th>Total balls</th>
<th>Mean collection rate</th>
<th>Mean No. of cones lifted after last collection</th>
<th>Probability of choosing</th>
<th>Probability of choosing</th>
</tr>
</thead>
<tbody>
<tr>
<td>ToLo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of moves</td>
<td>.36*</td>
<td>.4</td>
<td>.1</td>
<td>-.44*</td>
<td>.34*</td>
<td>.0</td>
<td>-.33*</td>
</tr>
<tr>
<td>Correct sequences</td>
<td>-.05</td>
<td>-.48*</td>
<td>-.29</td>
<td>.23</td>
<td>-.49*</td>
<td>-.17</td>
<td>.4</td>
</tr>
<tr>
<td>Rule violations</td>
<td>.16</td>
<td>.07</td>
<td>-.36*</td>
<td>-.2</td>
<td>.2</td>
<td>-.17</td>
<td>-.31*</td>
</tr>
<tr>
<td>Time violations</td>
<td>.41*</td>
<td>.17</td>
<td>-.33*</td>
<td>-.44*</td>
<td>.0</td>
<td>-.05</td>
<td>-.27</td>
</tr>
<tr>
<td>Latency to first move</td>
<td>.37*</td>
<td>.06</td>
<td>-.25</td>
<td>-.25</td>
<td>-.06</td>
<td>0</td>
<td>-.01</td>
</tr>
<tr>
<td>Total time</td>
<td>.36*</td>
<td>.19</td>
<td>-.36*</td>
<td>-.45*</td>
<td>.14</td>
<td>-.06</td>
<td>-.35*</td>
</tr>
<tr>
<td>Execution time</td>
<td>.37*</td>
<td>.17</td>
<td>-.35*</td>
<td>-.42*</td>
<td>.1</td>
<td>-.05</td>
<td>-.32*</td>
</tr>
</tbody>
</table>

Note: Asterisks mark significant ($p < .05$) values.

cidentally turned off the camera while searching. Thus, the final sample included 48 participants (mean age = 9 years [SD = 2.91], 85% male, mean IQ = 97.5 [SD = 8.3]).

Regarding the association between the scores of the ToLo and the patchy version of the BSFT, we found several significant correlations (Table 2). Overall, the $r_s$ values indicate that participants with a poor performance on the ToLo also produced a less efficient searching performance on the BSFT (e.g., more moves on the ToLo correlated with a lower rate of ball collections and a lower probability of choosing a close cone after a collection). Table 2 also shows that some measures indicative of good performance on the BSFT (e.g., the lower probability of choosing a close cone after an empty cone or the slope of collection rate) were not associated with performance on the ToLo.

DISCUSSION AND CONCLUSION

Regarding our main aim, which was to relate performance on the ToLo and BSFT to start piecing together the possible neuropsychological correlates suggested by the comparison between different types of tasks – those taking place over a desktop, common in office space evaluations, and those occurring outside traditional evaluation settings – we can refer to finding significant correlations between the BSFT and the ToLo, but also to those measures without a significant association.

The similarities between the BSFT and ToLo are several. Both consist of spatial problem solving tasks that involve planning and working memory to solve the problem in an efficient manner. To perform well on either of these tasks, the participant must be able to comprehend, although not necessarily consciously, the underlying “trick” needed to solve the task – in the case of the ToLo, success can be related to insight that to solve the task, often counter-intuitive moves must be performed (e.g., removing a ball from the correct peg in order to remove the ball from beneath). Similarly, as participants visit more patches in the BSFT, it is advantageous for them to recognize that the balls are arranged as a cluster. From the description of the performance variables, one can readily appreciate how re-visiting patches in the BSFT or how making many moves in the attempt to solve the ToLo decreases efficiency. The fact that several performance measures on the BSFT showed significant correlations when compared to performance on the ToLo suggests that both tasks may provide indirect measures of similar cognitive abilities. For instance, to perform well on the BSFT, participants need to work out the spatial array of the targets and thus increase their overall collection rate (attention and problem solving), leave a patch after suspecting most of the balls have been collected (working memory), or search for balls under nearby cones in the expectation of clusters (pattern recognition, planning).

The number of significant correlations between this latest version of the BSFT and the ToLo was larger than those for previous versions of the BSFT (Rosetti et al., 2018). Previous comparisons mainly highlighted the association between the number of violations on the ToLo and inefficient search patterns on the BSFT such as returning to previously visited cones. The explanation for this similarity appeared to be rooted in a lack of inhibition or attention, as rule violations are mainly encountered in clinical samples of patients with ADHD (Culbertson & Zillmer, 1998). Looking at the current results, we can say that poor performance on either of these tasks could be linked to similar aspects to those highlighted by Culbertson and Zillmer (1998) regarding cognitive impairments associated with lack of planning and poor problem solving. The current sample was composed of children and adolescents diagnosed with ADHD, and thus the observed correlations may be reflecting similar deficits, particularly those relating to the poor performance of a clinical sample. This interpretation could be further strengthened by the inclusion of healthy controls to assess the rel-
ative performance of patients with ADHD on the “patchy” BSFT; testing of an equivalent healthy control group is currently in progress. For instance, that we did not observe a correlation between measures on the ToLo and the slope of the collection rate may be related to the fact that only a few participants showed a large, positive slope (indicative of good information reception and application). This (lack of) association could be different in a healthy population where participants may be quicker to detect the clustered within-patch distribution of targets as they progress across successive patches. The addition of a control group could thus help identify the performance measures that provide the best profiling values of BSFT as an assessment tool.

Here it is important to remark that the purpose behind the continuing development of the BSFT is not to perfectly match it to other forms of assessment but rather to better understand a new and hopefully complimentary tool, which is based on many different underlying principles to various previous, mainly table-top, tests. For instance, while the ToLo limits, by design, the way in which participants can best solve the task, BSFT participants can adequately solve (or fail to solve) the task in various ways. For example, thoroughly searching a patch to deplete it versus leaving after inspecting only a few cones, hoping to collect more targets by visiting more patches, may result in a similar number of collected targets although perhaps with different energetic expenditure (Pacheco-Cobos, Rosetti, Cuatianquiz, & Hudson, 2010). Examining the diversity in searching strategies could be useful in detecting clinically distinctive subgroups: participants who loose efficiency by spending much time on the same patch versus participants who loose efficiency by moving quickly from patch to patch, including returning to previously visited patches. Furthermore, the task can be applied to a wide range of ages – simple modifications can be made to adjust the degree of difficulty to the motor and cognitive abilities of even very young children. Finally, the BSFT involves a large degree of sensorimotor feedback: moving (often running) between potential targets, lifting the cones and collecting the balls, and carrying collections around for the duration of the task. This feedback may help maintain motivation and provide a positive feeling of achievement even in the case of poor performance. Imperfect correlations in the present study between scores on the ToLo and BSFT support the need for complementary tests for the evaluation and better understanding of psychiatric disorders such as ADHD. Such differences with more established test methods are to be expected as the effort to provide more ecologically valid measures of cognitive function proceeds.

Funding

This research was supported by grants from the Consejo Nacional de Ciencia y Tecnología (CONACyT, No. 162043), by the Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PIAPI, No. 1A202617), and by institutional funding from the Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México.

Conflict of interests

Authors declare no conflict of interests.

Acknowledgments

We thank all participants for their enthusiastic collaboration as well as doctor Gina Chapa and doctor Alejandra Hernández and psychology intern Alejandro Angulo who assisted with recruiting participants, performing assessment procedures and coding the obtained results.

REFERENCES


Neuropsychological correlates of search behavior in ADHD

adolescents diagnosed with attention-deficit/hyperactivity disorder: Compari-
son with two standard neuropsychological methods. Journal of Clinical and
Experimental Neuropsychology, 49(6), 576-585.
performance in a searching task. Journal of Environmental Psychology, 49, 86-
95.
Rosetti, M. F., Ulloa, R. E., Vargas-Vargas, I. L., Reyes-Zamorano, E., Palacios-Cruz,
with ADHD on the Ball-Search Field Task. Scientific Reports, 6, 19664.
Sbordone, R. J. (1996). Ecological validity: some critical issues for the neuropsycho-
logist. In R. J. Sbordone & C. J. Long (Eds.), Ecological validity of neuropsycho-
logical testing (pp. 15-41). Boca Raton: St Lucie Press.
Neuropsychological function in adults with attention-deficit hyperactivity dis-
Shallice, T. (1982). Specific impairments of planning. Philosophical Transactions of
the Royal Society of London Series B: Biological Sciences, 299(1089), 199-209.
Sheehan, D. V., Sheehan, K. H., Shytle, R. D., Janavs, J., Bannon, Y., Rogers, J. E., …
& Wilkinson, B. (2010). Reliability and validity of the mini international neuro-
psychiatric interview for children and adolescents (MINI-KID). The Journal of
Clinical Psychiatry, 71(3), 313-326.
Wechsler, D., Kaplan, E., Fein, D., Kramer, J., Morris, R., Delis, D., & Maelender,
Validity of the executive function theory of attention-deficit/hyperactivity disor-
der: A meta-analytic review. Biological Psychiatry, 57(11), 1336-1346.