

# Resilience of the Tower Test to Response Bias

Thomas M. Dunn\*, Amy Rhodes, Shanda D. Crowder

Department of Psychological Sciences, University of Northern Colorado, Greeley, CO 80639, USA

\*Corresponding Author: [Thomas.dunn@unco.edu](mailto:Thomas.dunn@unco.edu)

Copyright © 2013 Horizon Research Publishing All rights reserved.

**Abstract** Neurocognitive response bias is a concern of clinical neuropsychologists, as accurate assessment is not possible if the patient being tested is not putting forth maximum effort during testing. Despite decades of research in this area, very little study has specifically examined the resilience of neuropsychological tests to incomplete effort. When college students asked to feign cognitive deficits are assessed on the Tower Test from the Delis – Kaplan Executive Function System (D-KEFS), they perform similarly to control participants (asked to do their best) on several Tower Test scales and significantly better than those with known clinical deficits. These results suggest that the Tower Test may have some resiliency to neurocognitive response bias.

**Keywords** effort, malingering/symptom validity testing, neuropsychological assessment, tower test, TOMM, D-KEFS

---

## 1. Introduction

In recent years, there has been significant interest in the accuracy of neuropsychological assessment with patients who are not putting forth maximum effort during testing.[1,2] Incomplete effort is of major concern to clinical neuropsychologists, as the validity of cognitive assessment is dubious if the patient is not putting forth their maximum effort.[3] Patients undergoing neuropsychological assessment may be subject to a number of influences that prevent them from performing at their highest possible level of functioning and therefore have biased responding. Some patients may simply be fatigued after a long day of testing and stop working as hard as they are truly capable of doing.[4] Misunderstanding test directions, lack of interest in the assessment process, and feeling overwhelmed or frustrated by the assessment process may also result in response bias during neuropsychological testing.[5] Patients are also known to manufacture cognitive symptoms in order to achieve secondary gain,[6] biasing their responses in a manner that makes them present with deficits that do not really exist.[7] Finally, there has been evidence that even patients with true neurocognitive deficits may also bias their

responses, particularly when the outcome of neuropsychological evaluation may involve compensation.[8]

### 1.1. Detecting Response Bias

The detection of response bias and outright malingering is obviously a concern of neuropsychologists. This concern is especially salient during forensic assessments. Ardolf et al.[9] suggest that over three-quarters of assessments in forensic settings assessing neurocognitive dysfunction involve some degree of response bias, while Youngjohn, et al.[10] speculate that up to half of worker compensation claims may involve feigned symptoms. Indeed, a survey of expert practices in clinical neuropsychology found that almost all clinicians (95%) regularly comment upon the validity of their assessment when reporting test results.[11]

Broadly, assessing response bias falls into three major categories.[3,12,13] First, neuropsychologists may use clinical acumen, considering inconsistencies in test data, noting discrepancies between mechanism and severity of injury and test scores, or finding contradictory statements during the clinical interview. Second, there has been a proliferation of commercially available standardized tests that assess response bias.[2,5] Such instruments are popular because they provide objective data regarding effort and can be included when the clinician has concerns regarding a patient's presentation. Finally, a number of standard instruments in a neuropsychological test battery can be interpreted to consider whether responses may have been biased,[14,15] or have their own validity scales, such as the MMPI-2.[16] Such an approach can be valuable when a standard fixed battery does not contain a measure assessing effort.

The use of standard neuropsychological instruments when assessing for response bias also has other advantages. Some common effort tests, such as the Rey 15 Item Test,[17] are quite simple and may be too transparent (and therefore ineffective) for a sophisticated patient who is deliberately trying to skew test data. Traditional cognitive measures have an additional advantage of being less transparent and do not require lengthening the test battery by adding another instrument. Such an approach also allows a clinician to return to test data long after the assessment has been

completed if later made aware of suspicions that the patient may not have been forthcoming about their presentation. Examples of cognitive tests that have been identified as being useful in the detection of response bias include the Digit Span subtest of the Wechsler Adult Intelligence Test – III,[18] the Wechsler Memory Test – III,[19] the Trail Making Test (parts A and B),[20] and the Seashore Rhythm Test.[21]

## 1.2. Instruments Resilient to Response Bias

Despite a literature of standard neuropsychological tests being used to detect biased responding, there is a paucity of research regarding measures that provide robust results despite incomplete effort and thus may be *resilient* to response bias. That is, results obtained from such an instrument are likely to yield a relatively accurate portrayal of the test taker's true performance despite incomplete effort. When considering various instruments that may be resilient to response bias, we speculated a tower test (e.g. Tower of Hanoi or Tower of London) may have properties suitable for resilience to response bias. Tower tests, common measures of executive functioning,[22-24] are disk transfer tasks that have simple rules and are relatively short to administer. Short tasks are less likely to cause fatigue in test takers and simple rules reduce the possibility of biased responding due to misunderstanding test instructions. Tower tests are also believed to be mastered largely relying on procedural memory.[25,26] This is significant, as procedural memory is thought to be an implicit process and outside consciousness (see Squire[27] for a discussion). When the Tower of Hanoi, for example, is administered serially, participants typically adopt more efficient methods to solve the puzzle and improve their performance.[28] Further, even amnesic patients tend to improve their performance over repeated administrations of the Tower of Hanoi,[29,30] as non-declarative procedural memory seems to be the system underpinning learning on this task.[31,32] Therefore, over repeated trials, it may be difficult for subjects who are trying to feign impairment to continue a consistently poor performance.

In order to study whether tower tests may be resilient to response bias, we selected the Tower Test subtest of the Delis – Kaplan Executive Function System (D-KEFS) test battery, as it has been shown to have robust psychometric properties and utility in clinical assessment.[33-35]. The Tower Test is based on the Tower of Hanoi (see Figure 1), the disk transfer task that was invented by the French mathematician Edouard Lucas in 1883 as a mathematical puzzle.[36] Test takers are asked to move different sized disks to one of three pegs of varying sizes. The Tower Test subtest of the D-KEFS employs a paradigm where the patient is administered multiple trials of tower tasks, thereby allowing for more efficient test taking strategies to develop. By using this instrument, we are able to make experimental hypotheses.



Figure 1. Tower of Hanoi

## 1.3. Hypotheses

- 1: Analog participants asked to fake a head injury would perform similarly on the Tower Test to those asked to do their best
- 2: Analog participants would achieve Tower Test scores significantly better than those found in the impaired range of functioning
- 3: Both analog participants and control subjects would perform differently than a clinical group composed of those with known neurological deficits.

## 2. Method

### 2.1. Participants

Non-clinical participants were recruited from an introductory psychology participant pool after IRB approval was granted. Students who completed the study received credit for their participation in research. Fifty-six participants (38 females) were recruited to be part of the non-clinical sample. These participants were randomly assigned to be controls ( $n=27$ ) or simulators ( $n=29$ ). Participants reporting a previous head injury ( $n=3$ ) were not included in this study, providing a sample thought to be free from overt neurocognitive pathology.

A third data set included clinical data published by Yochim et al.[37] and used with the permission of the senior author via E-mail with B. Yochim, PhD in July 2011. This clinical group comprised 12 individuals with known prefrontal cortex lesions and is well described by Yochim et al. [37]. Briefly, the average (and SD) age of this clinical groups was 63 years (12.7) and the mean years of education was 14.2 (2.6). The etiology of the lesion was cerebral vascular accident for ten of their participants; the remaining two had intracranial space occupying lesions.

Yochim et al.[37] also provided guidance regarding sample size to achieve sufficient power (.80) with an alpha level of .05. The average effect size reported in their results ( $\eta^2 = .29$ ) was used to calculate non-centrality parameter  $\lambda$  of 11.06 ( $df = 2, 24$ ) suggesting that a total sample size of 30 would be sufficiently powered.

### 2.2. Materials

All participants were given the Tower Test from the

D-KEFS.[33] Non-clinical participants also took the Test of Memory Malingering (TOMM)[39] and the Vocabulary subset from the WAIS-IV.[44]Both non-clinical groups also filled out a demographics form and answered a brief, post-experimental questionnaire.

As mentioned previously, the D-KEFS Tower Test is based on the Tower of Hanoi and is a 9-item test that measures planning and problem solving abilities. It uses a board with three equal length vertical pegs and five colored disks of various sizes from small to large (see Figure 2). The participant is asked to move the colored disks, in the fewest moves possible, from a predetermined starting position to a specified ending position, or arrangement, displayed by the examiner. The minimum number of moves for each item ranges from 1 to 26, increasing in complexity as the task progresses. Common to disk transfer tasks, there are two rules that the examinee must follow during this task: the participant is allowed to move only one disk at a time and a larger disk may never be placed over a smaller disk.



**Figure 2.** Tower Test from the D-KEFS

Performance on the Tower Test is measured in several ways. A global indicator of performance is given via Total Achievement Score that is earned based on number of moves, time to completion, and number of towers completed. Low scores are associated with deficits in executive functioning. More specific optional process scores are also reported. Among these optional scores are Move Accuracy Ratio, the total number of moves across all items administered divided by the total number of minimum moves required across all

items. This score allows the examiner to determine what strategies the examinee used while constructing the towers. A Time-Per-Move Ratio score provides the average time an examinee takes for each move. High scores can indicate deficits in impulse control while low scores may indicate poor initiation. Finally, Total Rule Violations account for the total number of rule violations made throughout administration of the Tower Test. Rule violations are unusual and may indicate a failure to maintain a cognitive set.[33]

In order to compare the effort put forth by the non-clinical groups, a second measure, the TOMM, was also administered to the analog participants. It is a measure of effort employing a forced-choice visual recognition test consisting of 50 black-and-white drawings. The test is composed of three components: Trial 1, Trial 2 and a Retention trial. During the study phase, the participant views each drawing for three seconds with an interval of one second between drawings. After the 50 black-and-white drawings have been presented, they are then shown again, one-at-a-time, to the participant along with a distractor drawing. The participant then indicates which of the two drawings was shown during the study phase (Trial 1). During Trial 2, the task is repeated. After Trials 1 and 2, the Retention Trial is administered, beginning approximately 15 minutes after the end of Trial 2. The retention task, similar to Trials 1 and 2, is administered with the 50 earlier drawings paired with 50 distractor items with participants instructed to identify the drawing shown during the study phase. The TOMM has been shown to have robust psychometric properties.[38,39]

The Vocabulary subtest from the WAIS-IV was used as a rough proxy of intellectual functioning.[40] Like the TOMM, it was used to examine whether differences in intellectual functioning could account for differences between groups. The demographics form collected basic data such as age and history of head injury or learning disability.

### 2.3. Procedure

Recruited non-clinical participants were first met by an “experimenter” who gained informed consent, collected demographic data, administered the Vocabulary subtest and then randomly assigned participants to either a brain injury simulation group or a control group. The brain injury simulation group (“Simulator Group”) was intended to be an analog to individuals who present for neuropsychological testing and who do not put forth maximum effort during testing. In order to create a group of test takers who would “fake bad,” we created a scenario where we asked them to pretend that they had been in a fictitious car accident. We further asked them to pretend that they were experiencing symptoms like difficulty concentrating. The control group (“Control Group”) was asked to do their best. Specific instructions to the groups are found in the Appendix. Participants were then introduced to a blind “examiner” who

administered the TOMM and the D-KEFS Tower Test. The Tower Test was used as an intervening task between Trial 2 and the Retention Trial of the TOMM. Once they had completed the testing, participants were returned to the “examiner” who completed a debriefing.

Data analysis was achieved by using independent t-tests to compare differences between the two non-clinical groups on demographic variables, the Vocabulary subtest and TOMM performance. Both non-clinical groups were compared to each other and the Yochim et al.[37] clinical group using a one-way ANOVA and Fisher’s Least Significant Differences to perform follow-up comparisons. Finally, a single sample t-test was used to compare scores of the Simulator group to D-KEFS Tower Test norms.

### 3. Results

#### 3.1. Demographic Data

Table 1 shows relevant demographics for the non-clinical participants. Both the Simulator and Control groups were the same age and were predominantly women. Also reported are the standard scores from the Vocabulary subtest for both the Control and Simulator groups. Please note that these scores are nearly identical and an independent group t-test confirms no significant differences.

#### 3.2. Tower Test Data

All three (Control, Simulator and Clinical) groups were then compared on their performance on the Tower Test (see Table 2). A One-way ANOVA was used for significance

testing and follow-up tests were completed (when necessary) using Fischer’s Least Significant Difference (LSD). Effect sizes ( $\eta^2$ ) were also computed and can be found in Table 3. As previously mentioned, The Tower Test gives several scores regarding performance. When evaluated on the Total Achievement Score (an overall indicator of performance) the Clinical Group had the worst performance between these groups and performed significantly worse the Control Group. However, the Simulator Group was not significantly different than either the Clinical or Control Group. Both non-clinical groups had a mean Total Achievement Scores in the average range, while Clinical Group scored in impaired range.

On the Mean Time-Per-Move Ratio, the Clinical Group took more than six seconds per move, while the two non-clinical groups took about three seconds. The normative sample for the Tower Test had a Time-Per-Move Ratio of 3.6 seconds. While there were statistically significant differences between all three groups, the means of the non-clinical groups were roughly half a second apart and very similar to the normative sample for the instrument. There were no significant differences on the Move Accuracy Ratio measure, indicating that all three groups tended to employ effective strategies when constructing towers.

Finally, there was a marked discrepancy between Total Rule Violations between the clinical and two non-clinical groups. The Clinical Group had a mean of four violations versus about one violation for the other groups, a significant difference. The Simulator and Control Group were not significantly different and both scores of about 1.00 are consistent with the performance of the normative sample.[33]

**Table 1.** Age, gender and Vocabulary score of non-clinical groups and results of independent samples t –test with effect sizes

Demographic	Control Group (n = 27)	Simulators (n = 29)	t-test (54) =	p	d
Female Gender	19 (70%)	19 (65%)	N/A		
Mean (SD) Age in years	19.07 (2.04)	19.83 (4.65)	-.78	= .44	-.21
Mean (SD) Vocabulary Standard Score	11.29 (1.61)	11.76 (4.53)	-.50	= -.20	-.13

**Table 2.** Tower Test scores by group and results of One-Way ANOVA with effect sizes

Mean (SD) Tower Score	Yochim et al. Clinical Group (n = 12)	Non-Clinical Control Group (n= 27)	Non-Clinical Simulators (n = 29)	F (2,65) =	p	$\eta^2$
Total Achievement Score	13.50 (4.85)	18.11 (3.40) <sup>1</sup>	16.28 (4.37)	5.33	= .007	.14
Time-Per-Move Ratio	6.04 (2.19) <sup>2</sup>	2.86 (1.11) <sup>2</sup>	3.27 (1.22) <sup>2</sup>	22.97	< .001	.41
Move Accuracy Ratio	1.52 (.50)	1.34 (.16)	1.42 (.49)	.90	= .41	.03
Total Rule Violations	4.33 (3.65) <sup>3</sup>	.70 (1.98)	1.38 (3.75)	5.71	= .005	.15

Note. Fisher’s Least Significant Difference follow-up comparisons ( $p < .05$ ):

<sup>1</sup>Control Group scored significantly different than the Clinical Group.

<sup>2</sup>All groups are significantly different.

<sup>3</sup>Clinical Group scored significantly different than both Non-Clinical groups. There was no difference between the Non-Clinical Groups.

**Table 3.** Tower Test raw scores of the Simulator Group and impaired Tower Test values and single sample t-test results

Tower Score	Normative Raw Score Indicating Impaired Functioning <sup>1</sup>	Mean (SD) Raw Score Non-Clinical Simulators (n = 29)	t (28) =	p	d
Total Achievement Score	13	16.28 (4.37)	4.04	< .001	1.53
Time-Per-Move Ratio	4.7	3.27 (1.22)	-3.99	< .001	1.50
Move Accuracy Ratio	2.15	1.42 (.49)	-13.86	< .001	5.24
Total Rule Violations	2 <sup>2</sup>	.70 (1.98)	-3.42	< .001	1.29

<sup>1</sup>Raw score indicating impaired functioning was derived from Tower Test norms. It was calculated by finding the raw score that corresponded with a scaled score of 7 for the age group 19 years (the mean age for the Simulator Group). When a range of scores was given, the midpoint for the range was used.

<sup>2</sup>The Tower Test does not provide scaled score equivalents for Total Rule Violations, but two or more rule violations has been found to suggest impairment (Yochim et al., 2009)

**Table 4.** Test of Malingered Memory (TOMM) scores by group and results of independent samples t-test with effect size

Mean (SD) TOMM Scores by Trial	Control Group (n = 27)	Simulators (n = 29)	t test t (54) =	p	d
Trial 1	47.85 (5.05)	28.83 (9.75)	9.06	< .001	2.45
Trial 2	48.85 (5.97)	29.62 (12.74)	7.13	< .001	1.93
Retention	48.78 (.28)	29.24 (13.45)	6.00	< .001	2.05

Note. The raw TOMM scores of number correct out of 50 are reported.

### 3.3. Comparison of Simulator Group to Tower Test Norms

It was also desirable to compare the Simulator Group to Tower Test norms. This was done in two ways. First, most Tower Test scores have scaled-score equivalents based on age.[33] Using the average age of Simulator Group (19), its raw scores could be converted into scaled-score equivalents. The Simulator Total Achievement scaled-score was 9, the Time-Per-Move scaled-score was 10, and the Move Accuracy Ratio was 11 (Total Rule Violations do not have scaled-score equivalents.[33]Secondly, we wished to compare the Simulator Group’s mean performance to what would be considered impaired scores when following D-KEFS Tower Test norms. For this analysis, we compared mean raw scores from the Simulator Group to those corresponding with a scaled score 7, as this suggests some degree of impairment (please see Table 3). Significance testing was achieved using a single sample t test (with the raw score corresponding with a scaled score of 7 as the test value) and is also shown in Table 3. There was a marked and statistically significant difference between the Simulator Group mean performance and scores suggesting impaired functioning for Total Achievement Score, Time-Per-Move and Move Accuracy Ratio. As cumulative percentages are reported and not scaled scores in the normative data, [33] we used a value of 2 total rule violations, as other investigators have suggested scores of two or higher suggest impairment.[37] As was the case in previous analyses, the Simulator Group scored significantly lower than 2 rule violations.

### 3.4. Check to Verify Participants Followed Experimental Directions

As an experimental check, both non-clinical groups also

were evaluated on the TOMM (Table 4). Indeed, the Simulator Group performed very badly and significantly different than the Control Group. The mean score for the Simulators is well into the range of incomplete effort and far below what is typically expected for even those suffering from neurological impairment.[39]

## 4. Discussion

### 4.1. Study Goals and Hypotheses

The goal of the study was to determine whether a tower test demonstrated resilience to response bias. We selected the Tower Test from the D-KEFS, a standardized version of the Tower of Hanoi to place in an experimental paradigm. Previous research has suggested that most test takers improve their performance across trials, even with significant memory deficits.[28-30] We hypothesized that non-clinical participants asked to simulate head injury (Simulators) would perform comparably to a Control Group. Further, we hypothesized that both groups would perform better than a clinical sample.

### 4.2. Findings

Perhaps the most noteworthy finding of this study is the performance of the Simulator Group on the Tower Test. Generally, when a group of neurologically intact individuals are asked to simulate head injury during neuropsychological testing, they perform very poorly, often performing worse than those with true neuropsychological impairment.[6,13,14] That groups asked to put forth incomplete effort generally produce markedly impaired scores is not surprising, as maximum effort is the typical

expectation for valid neuropsychological impairment.[1,3] However, when evaluating performance on the Tower Test, the Simulator Group scored in the intact range for all Tower measures and performed comparably to a control group asked to do their best.

Specifically, of the various performance scores on the Tower Test, Simulators did not perform significantly different than the Control Group on the Total Achievement Score, Total Error Violations, or Move Accuracy Ratio scores. While they did show a statistically longer average Move-Per-Time Ratio (3.27 seconds), they were less than a half-second longer than the Control Group (2.86 seconds) and still considerably faster than the Clinical Group (6.04 seconds). Further, the mean performance of the Simulators did not score in a manner that suggested neuropsychological impairment and was quite consistent with the normative sample for the D-KEFS. This is an interesting finding.

Given that we believed the Simulator Group was, as a whole, attempting to feign cognitive impairment while taking the Tower Test, it was also desirable to compare findings of the Simulator Group in a clinical context. This was done in two ways. First, following D-KEFS guidelines for scores suggesting neuropsychological impairment, the mean score for those in the Simulator Group does not fall in the impaired range on any of the four measures. Second, the Simulator Group was compared to the Yochim et al.[37]published Clinical Group. This comparison revealed mixed results.

Despite a Total Achievement Score of 16 for the Simulators and 13 for the Clinical Group, this difference was not statistically significant. While this is noteworthy, Delis et al. (2001) recommends that the Tower Test's optional process scores be used to help determine areas of specific deficiency when the Total Achievement Score suggests impairment.[33] When these scores are compared between the Clinical Group and Simulator Group, there are important differences. First, the Simulator Group scored significantly better than the Clinical Group on Time-Per-Move Ratio and on Total Rule Violations. While the groups scored the same on the Move Accuracy Ratio, their scores were not suggestive of impairment, nor were they significantly different than the Control Group. Additionally Yochim et al.[37] cogently argue that rule violations on the Tower Test are extremely valuable when considering executive function impairment.[37] They note a high level of sensitivity and 100% specificity in frontal lobe lesion detection by examining Total Rule Violations. Given that the Simulator Group performed no differently than the Control Group and significantly better than the Clinical Group regarding rule violations suggests that the Tower Test results are robust even when persons are trying to feign impairment.

#### 4.3. Alternative Explanations

When considering alternative explanations that may account for these findings, we explored whether the "Fake Bad" and Control Group groups had similar performances on

the Tower Test because the Simulator Group did not understand the experimental directions to *not* do their best and therefore tried their hardest when asked to complete the neuropsychological tests. Rogers and Cavanaugh[41] regard this as the "simulation-malingering paradox," because participants are asked to comply with directions to fake-bad in order to provide information about people who refuse to follow test directions. To address this potential confound, we used a standard test of effort, the TOMM. If the Simulators did not understand the experimental instructions to fake head injury, we believe this tendency would have resulted in TOMM scores similar to the Control Group. This was not the case. Indeed, as the Tower Test was used as the intervening task between the TOMM's second trial and the retention trial, it is unlikely that the simulator group, who performed poorly across all trials of the TOMM, would follow directions to fake head injury with this task, while not following the same directions on the Tower Test. It is, therefore, reasonable to rule out inability to follow test directions as an explanation of these findings.

Another potential explanation for the finding that non-significant differences were found between groups is that the study was underpowered. Certainly, with small sample sizes, there is the possibility of committing a Type II error.[42] An *a priori* power analysis suggested adequate power with a sample size of about 30. Given that our total sample size was greater than 60 and that on several measures significant differences were detected, we do not believe insufficient power is the best explanation for our results.

Indeed, when considering the performance on the Tower Test of the Simulator Group in context of the normative data for the instrument, relative to a control group, and compared to a clinical group, the results tend to support our hypotheses. Neurologically intact individuals tend to perform well on the Tower Test despite their efforts to present as cognitively impaired. We suspect that this is due to procedural learning being somewhat outside the conscious control of the individual.[25,32] The D-KEFS' Tower Test paradigm of asking test takers to repeatedly solve tower problems of varying difficulty makes the notion of procedural (and unconscious) learning having a strong role in these results quite plausible.

#### 4.4. Interest to Clinicians

That a neuropsychological instrument has relative resilience to response bias may have important ramifications. For clinicians, these findings suggest that patient performance on the Tower Test is likely to be robust despite concerns regarding response bias, particularly when interpreting Total Rule Violations. On one hand, this may provide the clinician with valid assessment, to some degree, of executive functioning (such as response inhibition, impulse control and rule learning) when other frontal lobe measures may have questionable utility because the patient may be biasing their responding. On the other hand, clinicians should be aware that this resilience exists.

Particularly when interpreting test data from an entire battery, a believable profile on the Tower Test should not necessarily lend credibility to the notion that the patient is putting forth maximum effort.

#### 4.4. Interest to Researchers

We believe these findings are also important for researchers whose interest is response bias. As we believe the procedural learning component of the Tower Test is what accounts for its relative resistance to response bias, then other instruments that also tap non-declarative memory should have similar results. Mirror tracing, for example, may be an example of another such instrument with properties that warrant further study to examine its resilience to response bias.[43]

#### 4.5. Study Limitations

Despite these interesting results, there are several limitations to this study. Among them is the composition of the Clinical Group. Ideally, it would have been desirable to find a clinical group that perfectly mirrored our non-clinical control group (college aged patients whose neurological impairment is secondary to traumatic brain injury). We also do not have TOMM data for the Clinical Group; nor can we be absolutely certain that every member of this group was completely free from a motive to bias their responding. While unlikely, it is possible that comparison to another clinical group whose members have been carefully vetted to assure no reason for secondary gain may be different. Finally, as is the case in any study that uses analog participants, it is possible that people with an authentic motivation to bias their responding would perform differently on the Tower Test.

#### 4.6. Summary of Important Findings

Research into neurocognitive response bias tends to be centered on its detection. No research, that has been widely disseminated, has addressed whether there are neuropsychological instruments that are resilient to response bias. We placed the Tower Test from the D-KEFS into a common experimental paradigm to examine performance of analog participants' performance when asked to fake bad. Generally, the direction of incomplete effort will result in very low test scores on most neuropsychological measures. This was not the case, however, with the Tower Test. Indeed, a group asked to simulate head injury performed comparably to a control condition asked to their best. This is noteworthy as it suggests that some neuropsychological instruments may give valid results despite a test taker not putting forth maximum effort.

## Funding

No external funding was used for this study.

## Appendix

### Instructions for Simulator Group

For this study, we would like for you to pretend that you have been in a car accident. You were at an intersection; a car ran a stoplight and crashed into you. While no one was seriously injured, both vehicles sustained a sizable amount of damage. The other driver's car was totaled in the accident. You were not seriously injured, but you did require admission to a hospital. While seeming to recover fully, you notice at times you have some considerable aggravation, trouble concentrating, and have difficulty formulating ideas. Your lawyer suggests that you try to receive a \$50,000 settlement from an insurance company. This is only possible if you can prove you have suffered a head injury. Please do your best to fake having symptoms from a head injury. This would include poor concentration, trouble paying attention, difficulty formulating thoughts, and memory loss. In a moment you will be asked to take some neuropsychological tests. Be sure to take it as if you are the person in the story above who suffered a head injury. Do your best to be as convincing as possible. Please do not say anything to the person giving you the tests; please stay "in character". You can stop pretending to have a head injury when you leave the testing room.

*Do you have any questions before we begin?*

### Instructions for the Control (Do Best) Group

Thank you for agreeing to be in our study. In a few moments, you will be asked to take some memory tests. Please try as hard as you can and do the best job possible. At any time during the study, for any reason, should you feel uncomfortable, you may ask to leave the study. The stress you experience on a daily basis should not be exceeded as a result of being a part of this study.

*Do you have any questions before we begin?*

---

## REFERENCES

- [1] Brennan, AM, Gouvier WD. Are we honestly studying malingering? A profile and comparison of simulated and suspected malingerers. *ApplNeuropsychol.* 2006; 13: 1 -11. doi: 10.1207/s15324826an1301\_1
- [2] Sharland, MJ, Gfeller JD. A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Arch ClinNeuropsychol.*2007;22: 213 – 223. doi: 10.1016/j.acn.2006.12.004
- [3] HeilbronnerRL, Sweet JJ, Morgan JE, LarrabeeGJ, Millis, SR. American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological

- assessment of effort, response bias, and malingering. *ClinNeuropsychol.*2009; 23(7): 1093-1129. doi:10.1080/13854040903155063.
- [4] Utzl B, Graf P, Cosentino S. Exacting assessments: Do older adults fatigue more quickly? *J ClinExpNeuropsychol.* 2000; 22: 496 –507. doi: 10.1076/1380-3395(200008)22:4;1-0;FT496
- [5] Dunn TM. Current issues in response bias during neuropsychological assessment: Incomplete effort to malingering,199-217. In: J.R. Dupri (Ed.), *Focus on Neuropsychology Research*. New York: Nova Science Publishers; 2006.
- [6] Rogers RR. Current status of clinical methods, 391-410. In: R.R. Rogers (Ed.). *Clinical assessment of malingering and deception*(3rd ed.).New York: Guilford; 2008.
- [7] Slick DJ, Sherman EMS, Iverson GL. Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *ClinNeuropsychol.* 1999; 13: 545 – 561. doi: 10.1076/1385-4046(199911)13:04;1-Y;FT545
- [8] Yochim BP, Kane KD, Horning S, Pepin R. Malingering or expected deficits? A case of herpes simplex encephalitis. *Neurocase: The Neural Basis of Cognition.* 2010; 16: 451-460. doi: 10.1080/13554791003623334
- [9] Ardolf BR, Denney RL, Houston CM. Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *ClinNeuropsychol.* 2007; 21(6): 899-916. doi:10.1080/13825580600966391
- [10] Youngjohn JR, Burrows L, Erdal K. Brain damage or compensational neurosis? The controversial post-concussive syndrome. *ClinNeuropsychol.*1995; 9: 112 – 113. doi: 10.1080/13854049508401593
- [11] Slick DJ, Tan JE, Strauss EH, Hultsch, DF. Detecting malingering: A survey of experts' practices. *Arch ClinNeuropsychol.* 2004; 19: 465 – 473. doi: 10.1016/j.acn.2003.04.001
- [12] Iverson GL, Binder LM. Detecting exaggeration and malingering in neuropsychological assessment. *J Head Trauma Rehabil.* 2000; 15: 829 – 885.
- [13] Nies KJ Sweet JJ. Neuropsychological assessment and malingering: A critical review of past and present strategies. *Arch ClinNeuropsychol.* 1994; 9: 501 – 552. doi: 10.1016/0887-6177(94)90041-8
- [14] LarrabeeGJ. Detection of malingering using atypical performance patterns on standard neuropsychological tests. *ClinNeuropsychol.*2003;17: 410 – 425. doi: 10.1076/clin.17.3.410.18089
- [15] Van GorpWG, Humphrey LA, Kalechstein A, BrummVL, McMullen WJ, Stoddard M. Pachana, NA. How well do standard clinical neuropsychological tests identify malingering? A preliminary analysis. *J ClinExpNeuropsychol.*1999; 21: 245 –250. doi: 10.1076/clin.17.1.69.15624
- [16] Henry GK,HeilbronnerRL, Mittenberg W, Enders C. The Henry-Heilbronner Index: A 15-item empirically derived MMPI-2 subscale for identifying probable malingering in personal injury litigants and disability claimants. *ClinNeuropsychol.*2006; 20(4): 786-797. doi:10.1080/13854040500287749
- [17] Lezak MD. *Neuropsychological Assessment.* (3rded.). New York: Oxford;1995
- [18] Axelrod BN, Fichtenberg NL, Millis SR, Wertheimer JC. Detecting incomplete effort with digit span from the Wechsler Adult Intelligence Scale – 3rd edition. *ClinNeuropsychol.*2006;20: 513- 523. doi: 10.1080/13854040590967117
- [19] KillgoreWDS, DellaPietra L. Using the WMS-III to detect malingering: Empirical validation of the rarely missed index (RMI). *J ClinExpNeuropsychol.*2000; 4: 49 – 64. doi: 10.1076/jcen.22.6.761.960
- [20] Iverson GL, Lange RT, Green P, Franzen MD. Detecting exaggeration and malingering with the Trail Making Test. *ClinNeuropsychol.* 2002; 16:398 – 406. doi: 10.1076/clin.16.3.398.13861
- [21] Curtis KL, Greve KW, Brasseux R, Bianchini KJ. Criterion groups validation of The Seashore Rhythm Test and Speech Sounds Perception Test for the detection of malingering in traumatic brain injury. *ClinNeuropsychol.* 2010; 24(5): 882-897. doi:10.1080/13854041003762113
- [22] Simon HA. The functional equivalence of problem solving skills. *Cogn Psychol.* 1975; 7: 268 – 288. doi: 10.1016/0010-0285(75)90012-2
- [23] Sullivan, JR, Riccio CA, Castillo CL. Concurrent validity of the Tower Tasks as measures of executive function in adults: A meta-Analysis. *ApplNeuropsychol.* 2009; 16(1): 62-75. doi:10.1080/09084280802644243
- [24] Welsh MC, Huizinga M. The development and preliminary validation of the Tower of Hanoi-Revised. *Assessment.*2001; 8: 167 – 176. doi: 10.1177/107319110100800205
- [25] Beaunieux H, Hubert V, Witkowski T, Pitel A, Rossi S, Danion J, Desgranges B, Eustache F. Which processes are involved in cognitive procedural learning? *Memory.*2006; 14: 521 – 539. doi: 10.1080/09658210500477766
- [26] Davis HP, Klebe, KJ. A longitudinal study of the performance of the elderly and the young on the Tower of Hanoi Puzzle and Rey Recall. *Brain and Cogn.*2001; 46: 95 – 99. doi: 10.1006/brcg.2000.1269
- [27] Squire LR. Memory systems of the brain: A brief history and current perspective. *Neurobiol Learn Mem.* 2004; 82: 171 – 177. doi: 10.1016/j.nlm.2004.06.005
- [28] Xu Y, Corkin S. H.M. revisits the Tower of Hanoi puzzle. *Neuropsychology.* 2001; 15: 69 – 79. doi: 10.1037/0894-4105.15.1.69
- [29] Markowitsch, HJ, von CramonDY, Schuri U. Mnestic performance profile of a bilateral diencephalic infarct patient with preserved intelligence and severe amnesic disturbances. *J ClinExpNeuropsychol.*1993; 15: 627 – 652. doi: 10.1080/01688639308402586
- [30] Saint-Cyr, JA, Taylor AE, Lang AE. Procedural learning and neostriatal dysfunction in man. *Brain.* 1998; 111: 941 – 959. doi: 10.1093/brain/111.4.941
- [31] Graf P, Squire LR, Mandler G. The information that amnesic patients do not forget. *Learn MemCogn.* 1994; 10: 164 –



178.doi: 10.1037/0278-7393.10.1.164

- [32] Salmon DP, Butters, N. Neurobiology of skill and habit learning. *Curr Opin Neurobiol.* 1995; 5: 184 – 190. doi: 10.1016/0959-4388(95)80025-5
- [33] Delis D, Kaplan E, Kramer J. *Delis–Kaplan Executive Function System.* San Antonio, TX. The Psychological Corporation, Harcourt Brace & Co; 2001.
- [34] Homack S, Lee D, Riccio CA. Test review: Delis-Kaplan Executive Function System. *J ClinExpNeuropsychol.* 2005; 27: 599 – 609. doi: 10.1080/1380339049091844
- [35] Heled E, Hoofien D, Margalit D, Natovich R, Agranov E. The Delis-Kaplan Executive Function System Sorting Test as an evaluative tool for executive functions after severe traumatic brain injury: a comparative study. *J ClinExpNeuropsychol.* 2012; 34: 151-159. doi:10.1080/13803395.2011.625351.
- [36] Hinz AM. Pascal's Triangle and the Tower of Hanoi. *Am Math Mon.* 1992; 99: 538 – 544.
- [37] Yochim BP, Baldo JV, Kane KD, Delis DC. D-KEFS Tower Test performance in patients with lateral prefrontal cortex lesions: The importance of error monitoring. *J ClinExpNeuropsychol.* 2009; 31: 658-663. doi: 10.1080/13803390802448669
- [38] Teichner G, Wagner MT. The Test of Memory Malingering (TOMM): normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Arch ClinNeuropsychol;* 2004: 19, 455-464. doi: 10.1016/S0887-6177(03)00078-7.
- [39] Tombaugh TN. *Test of Memory Malingering (TOMM).* North Tonawanda, NY: Multi-Health Systems; 1996.
- [40] Wechsler D. *Wechsler Intelligence Scale for Adults—4th Edition (WAIS-IV®).* San Antonio, TX: Harcourt Assessment; 2008.
- [41] Rogers R, Cavanaugh JL. Nothing but the truth. . .: A reexamination of malingering. *Journal of Law and Psychiatry.* 1993; 11: 443 – 460.
- [42] Cohen J. *Statistical Power Analysis for the Behavioral Sciences (2nded.).* Hillsdale NJ: Erlbaum; 1988.
- [43] Gabrieli JDE, Corkin S, Michel SF, Growdon JH. Intact acquisition and long-term retention of mirror-tracing skill in Alzheimer's disease and in global amnesia. *BehavNeurosci.* 1993; 107: 899-910. doi: 10.1037/0735-7044.107.6.899.