Ursidae: The Undergraduate Research Journal at the University of Northern Colorado

Volume 1 | Number 2 Article 2

January 2012

The Delis-Kaplan Executive Functions System – Tower Test Resilience to Response Bias

Elizabeth Deborah Lowe

Abstract

Neuropsychological tools are used to gain more accurate insight about an individual's level of functioning (cognitive, behavioral, executive, etc.), and to make more exact diagnoses; therefore, valid neuropsychological tools are necessary for precise evaluation. Valid neuropsychological assessment relies upon the individual putting forth maximum effort during testing. While the literature is rich when describing methods of detecting incomplete effort, it is sparse when identifying instruments resistant to such response bias. The goal of this study was to determine whether or not effort affects performance on the D-KEFS Tower Test by comparing the results with the Test of Memory Malingering (a neuropsychological assessment designed to measure effort). Thirty-nine neurologically intact college students from a medium sized Rocky Mountain university introductory subject pool were asked to participate. The participants in the experimental group were given a vignette explaining that they had been in a car accident. The participants were then asked to pretend that they had suffered a brain injury and were having memory problems. The participants in the control group were asked to do their best. A blind examiner administered the D-KEFS Tower Test and The Test of Memory Malingering (TOMM) to both groups. Data analysis shows that there was a significant difference between the two groups' performance on the TOMM, but no significant difference between the scores on the Tower Test. These results suggest that the D-KEFS Tower Test is relatively resilient to incomplete effort.

Key Terms

malingering, response bias

Neuropsychological evaluation is an important factor in accurately assessing cognitive, behavioral, language, and executive dysfunction of an individual (Malik, Turner, & Sadler, 2007). The data obtained from neuropsychological evaluation provide insight that may lead to more accurate diagnoses. Neuropsychological assessments are useful tools; however, an important factor to keep in mind when interpreting the results is that these tools are sensitive to influences such as age, education, sex, and cultural background. All of these elements challenge the validity of the neuropsychological evaluation. Thus, they should be addressed before any conclusions are drawn.

Another major factor that can influence an individual's performance during neuropsychological evaluation – and, therefore, the validity of the results – is response bias. Response bias occurs when individuals do not put forth maximum effort during neuropsychological evaluation. Response bias skews the results so that they no longer accurately reflect the abilities of the individual. According to Dunn (2006), manifestations of response bias can be arranged on a spectrum. On one end, incidental response bias includes behaviors such as fatigue. These behaviors may occur because the test is tiring, or because individuals are distracted by environmental factors, such as noise or the temperature of the room. These behaviors are categorized by unintentional factors. The next integer on the continuum of response bias is the exaggeration of existing symptoms, or the reporting of symptoms that were once present but no longer exist. These behaviors are more intentional than the first cluster. Even more severe and intentional than the exaggeration of present or previous symptoms – residing on the extreme end of the response bias spectrum – is the deliberate fabrication of symptoms for personal gain, known as malingering. An example of this deliberate malingering would be an individual who was in an accident and feigns dysfunction in order to receive a significant insurance pay out. Malingering results in purposeful lack of effort; therefore, malingering skews the results of the test.

According to a survey conducted by Sharland and Gfeller (2007), this deliberate malingering occurs in 5-30% of cases, and additional cases of response bias occur equally as often. Their research suggests that some form of response bias (like fatigue or distraction) can occur in up to 60% of cases in neuropsychological testing. With such high rates, measures of detecting response bias are vital for accurate assessment of an individual's functioning.

There are two ways to detect response bias (Dunn, 2006). The first method is to use an assessment that is so simple and that most individuals, including those who are actually impaired, will easily perform well. Since the task is so simple, poor performance would be a reflection of response bias. The second strategy employed to assess response bias is to use an assessment known as a Symptom Validity Test (Dunn, 2006). These tests appear to measure memory while actually assessing effort. Individuals who put forth maximum effort on these tests, even those with cognitive deficits, will get all or almost all of the items correct; low scores (below 50%) reflect poor performance, which suggests that response bias has occurred. These assessments are also known as effort tests.

The major neuropsychological measures designed to assess effort include the Rey 15-Item Memory Test (Dunn, 2006), one of the oldest measures of response bias, and the Test of Memory Malingering (Tombaugh, 1996). The Test of Memory Malingering (TOMM) is considered to be "one of the most frequently used" and most accurate test used for detecting minimal effort (Sharland & Gfeller, 2007). The TOMM is disguised as a memory test, and malingering individuals tend to score very lowly. As the TOMM is simple enough that individuals can score 50% by chance, abnormally low scores are a reflection of deliberate malingering, not actual ability.

Much research has been conducted to determine what kind of tools can detect malingering, as well as other forms of response bias. However, the literature is sparse in regards to research that investigates whether there are instruments that are resistant to response bias in any form. If such instruments do exist, an individual may malinger or become fatigued, but his or her performance on the test would be no different than if maximum effort was put forth, under ideal testing conditions. Though the response bias is existent, the conclusions drawn from an individual's performance on the assessment can be considered an accurate reflection of the individual's functioning, unaffected by response bias.

One area to investigate is whether tasks that rely upon implicit memory are resilient to response bias. Research suggests that when a task involves implicit memory, an individual improves by simply doing it. If this is the case, people may not be able to fake poor performance on tasks involving implicit memory.

A neuropsychological test that relies upon implicit memory is the Delis-Kaplan Executive Function Systems Tower Test (TT), which is designed to assess several different executive functions, "including spatial planning, rule learning, inhibition of impulsive and perseverative responding, and the ability to establish and maintain the instructional set" (Delis, Kaplan, & Kramer, 2001), as well as to measure an individual's ability to problem solve (Chan, Chen, Cheung, Chen, & Cheung, 2004). The TT is composed of a series of nine items, each one progressively more difficult than the last. While completing each item, participants inadvertently gain knowledge which they use to solve the upcoming items.

Research indicates that traumatic brain injury (TBI) impairs abilities including problem solving, cognitive flexibility, and executive functioning (Leon-Carrion et al., 1998)—the very abilities that the TT assesses. Further research has been conducted to investigate how patients suffering from TBI perform on the TT. According to Chan et al. (2004), patients with TBI took longer to initiate the first move, and they also took longer to complete each problem than individuals without TBI. While research has been conducted to see how individuals with actual brain injuries perform on the TT, little research has explored whether performance on the TT is influenced by response bias.

Recently, Doyle and Dunn (2010) attempted to investigate this issue; they sought to determine if the D-KEFS TT was resilient to effort. In their study, Doyle and Dunn administered all nine items of the TT, then had each participant repeat two of the moderately difficult items. The expectation was that the performance of the control group (the participants who put forth maximum effort) would be consistent with the norms, and that their performance on the repeated items would improve. The expectation of the experimental group (the participants who were malingering) was that they would perform inconsistently with the norms, and that their performance would remain stagnant on the repeated items. The results were inconsistent with the hypotheses. Therefore, the results indicated that the TT was resilient to effort. Resilience to effort or malingering–a type of response bias–may imply that the measure is also resistant to other forms of response bias. However, the study lacked an objective assessment verifying the validity of the participant's performance. Doyle and Dunn (2010) only used a follow-up questionnaire asking the participants to rate themselves: how well did they think they followed the directions they had been given. Perception is subjective. Participants could think that they had effectively malingered when in reality they had not. This current study is designed to address the confound in Doyle and Dunn's (2010) research, using the TOMM to objectively assess whether or not participants were following instructions.

We hypothesized that the performance of the control group on the TT and TOMM would be consistent with the norms outlined in the Delis-Kaplan Executive Functions System and Test of Memory Malingering manuals. Performing at a normal level on the TOMM would be indicative that participants were putting forth maximum effort. We expected that the performance of the experimental group on the TT would be atypically poor, and that malingering would be represented by abnormally low scores on the total achievement score. We also expected that the experimental group would perform poorly on the TOMM, showing low scores on each of the trials, indicating that participants were not putting forth maximum effort, malingering instead. If the control group's performance is consistent with norms, and the experimental group's performance is abnormally low and significantly different than the control, this would indicate that the TT is indeed affected by malingering.

Method

Participants

In order to determine how well the Tower Test detects malingering in comparison to the TOMM, 48 intellectually sound, introductory psychology students at a Midwestern university were asked to complete both the D-KEFS Tower Test as well as the TOMM. (Refer to Table 1 for demographic information.) Both males and females were randomly assigned into the control or experimental groups. The participating students received credit in their psychology class as compensation. However, participating in this study was only one way to receive that credit; there were alternative opportunities to obtain the credit.

Table 1

Demographics

	n
Gender	
Female	30
Male	18
Total	48
Age	
17	2
18	26
19	13
20	5
21	2

Total	48

Measures

The D-KEFS Tower Test was given to both a control and an experimental group. The TT consists of a flat wooden board that holds three equidistant pegs, as well as eight wooden disks of increasing size. In each of the nine items, up to eight disks were arranged on the board in a predetermined configuration. The participants transferred the disks one at a time, placing only smaller disks on top of larger disks, to build a tower that replicated the tower in a picture. Each item has a set arrangement to begin with and a counterpart that the final tower must emulate.

Performance on the Tower Test was determined by a single score which was calculated using the scoring system provided by the D-KEFS manual. Each individual item received a score based on whether or not the tower was correct, how many moves the participant made, and how long the participant took to complete the item. Once the score for each item was determined, all nine numbers were added together. The sum was the overall achievement score for the participant's performance on the Tower Test.

To verify that the participants were following instructions, it was necessary to use an objective measure which would detect effort. In this study, we employed the TOMM. The TOMM is a pictorial memory test that consists of three trials: Trial 1, Trial 2, and Retention. Each trial is made up of 50 items. In the first trial, the participant was shown each of the 50 pictures for three seconds each. Once all the pictures were shown, the participants were shown a series of 50 recognition panels. Each panel contained one of the original pictures and one novel image. The participants identified which image they had seen before. The second trial was administered immediately after the first. The participants were shown the original 50 pictures a second time, in a different order. Again, the first series of panels was followed by a series of recognition panels, this time with different novel images. The participants identified which image they recognized. The second trial was followed by a 15-minute delay when the examiner administered the Tower Test to the participant. At the end of the delay, the retention trial was administered. This time, the original pictures were not presented. Instead, the retention trial was composed of only 50 recognition panels. Each panel held one of the original images as well as a new novel image, and participants identified which image they recognized as having seen before. Performance on the TOMM is represented by three scores: each score is the sum of correct responses in each trial.

The materials used for this study were not limited to the tests themselves. The supplemental material included the informed consent, a demographic questionnaire, the directions for the control group, the experimental vignette, as well as a follow-up questionnaire, and a debriefing form. Additionally, a quarter was used to randomly assign individuals to the control or experimental conditions. The protocol employed during this study consisted of forms for the TOMM (for the examiner to record which responses were correct and incorrect on each trial) as well as a form for the TT (for the examiner to record for each item, the length of time, number of moves, and whether the final product was correct or incorrect).

Procedure

In order to conduct this study, there were two different types of people helping to perform the study. The first was an experimenter who greeted the participants and briefed them. The second was an examiner who was blind to the true nature of the study and to which group each participant was assigned. The second examiner administered the two tests.

The experimenter asked the participants to read and sign the informed consent, then to complete a short questionnaire asking their age, year in school, major, and whether they had ever suffered a brain injury. While the participants completed the questionnaire, the experimenter flipped a coin to determine whether the participant would be in the control or experimental group. The experimenter told those who were randomly assigned to the control group that they would be taking two neuropsychological tests, then directed them to do their best on both. The experimenter read a vignette to those assigned to the experimental group asking participants to pretend that they had been in a car accident; if they could prove they had suffered a brain injury, the insurance company would pay a settlement of $50,000. The experimental participants were asked to feign symptoms of a brain injury, including memory problems and poor concentration while taking two neuropsychological tests. They were also asked not to discuss these instructions with the experimenter conducting the two tests so that the examiner collecting data could remain blind. All participants were given the opportunity to ask questions, and then taken to the testing room.

The lead experimenter left the participant with the blind examiner who began by conducting the first and second trials of the TOMM, followed by all items of the TT, and then the retention trial of the TOMM. Upon completion of the tests, the participant returned to the first experimenter who provided a follow-up questionnaire consisting of three questions: how well did you remember the directions, how motivated were you to follow the directions, and how well did you do as a participant? The participants were debriefed and were provided with contact information in case they had any follow-up questions or concerns.

Once the data were collected, the tests were scored so that the results could be analyzed and interpreted. An alpha level of .05 was set, and an independent samples t-test was computed in order to determine if there was a significant difference between the scores.

Results

Once all of the data was collected and scored for each of the 48 participants, it was analyzed. Of the 48 participants, the data for two participants were not used because they admitted to having suffered a head injury, leaving 46 data points to analyze. In the analysis, an independent samples t-test was run to see if there was a difference between the control and experimental groups' performance on the TT and on the TOMM. The hypothesis was that there would be a difference between the performances of the two groups on the TT, as well as between the groups' performances on the TOMM.

The t-test showed several different things. First, on the TT, there was no statistically significant difference ($t(44)=-0.48$, $p>.05$) between the experimental group (M=16.86, SD=2.35) and the control group (M=16.48, SD=2.93). Second, the two groups' performances on the first trials of the TOMM (control: M=48.48, SD=3.43; experimental: M=32.76, SD=10.72) were significantly different: $t(44)=6.95$, $p<0.00$. Third, when the performance of the control (M=49.20, SD=2.77) and experimental groups (M=35.24, SD=13.26) were compared on the second trial of the TOMM, the t-test showed that there was a statistically significant difference: $t(44)=5.144$, $p<0.00$. Finally, when the means of the groups' performances on the TOMM Retention Trial were compared (control: M=49.60, SD=1.61; experimental: M=32.70, SD=14.14), a statistically significant difference was found between groups: $t(43)=5.94$, $p<0.00$.

The results of the independent samples t-test are inconsistent with the hypothesis: the control and experimental groups performed differently on the TT, and there were marked differences between the groups' performances on each trial of the TOMM. The statistically significant differences between the groups on each trial of the TOMM suggest that the differences are not due to chance.

Discussion

This study sought to address the confound in Doyle and Dunn's (2010) research: to verify that the participants in the experimental group were indeed malingering during the administration of the TT by including the TOMM in order to determine whether or not the TT is resilient to response bias. Participants completed the TT and the TOMM; some were asked to put forth maximum effort on both, while others were asked to malinger (a form of response bias). We expected that there would be a statistically significant difference between the control and experimental groups' performance on the TOMM as well as the TT. The data analysis reflected that there was a statistically significant difference between the control and experimental groups' performance on the TOMM (supporting the hypothesis). However, contrary to the hypothesis, there was not a statistically significant difference between the control and experimental groups' performance on the TT. This shows that the participants in both groups were following instructions: those in the control group were trying their best, and those in the experimental group were malingering. Since the participants in the experimental group were malingering, and the results show that there was no difference between the two groups' performance on the TT, it has been concluded that the results suggest that the TT is resilient to effort.

Doyle and Dunn (2010) found that there was a statistically significant difference between the experimental and control groups' performance on the TT, which is inconsistent with the results of the present study. However, Doyle and Dunn also found that when the groups repeated two of the items, neither group improved on one of the items, and only the control group improved on the other. Despite the difference between the control and experimental groups' performance on the TT, both groups performed above the level that D-KEFS specified as indicative of impairment. These findings suggest that individuals could not fake traumatic brain injury on the TT. The findings of the present study are consistent with those of Doyle and Dunn

(2010), as both the control and experimental groups performed above the D-KEFS determined level of impairment.

The finding that the TT may be relatively resilient to intentional response bias may imply that it is also resilient to more incidental forms of response bias such as fatigue or exaggeration of symptoms. However, more research needs to be conducted in this area. Future research should be conducted to investigate whether the TT is resilient to other forms of response bias (fatigue and distraction). Future research should also explore whether other neuropsychological tools, such as the Wisconsin Card Sorting Task or Trail Making Task, are also resilient to response bias. Additionally, in the development of future neuropsychological assessments, it may be worthwhile to consider adding a component of implicit memory to the task in an attempt to design the test to be less influenced by response bias.

Acknowledgements

I would like to thank my family for their unconditional love and support. Your encouragement has given me the motivation to keep going when I haven't thought I could continue. I would also like to thank my friends for their support as well: having study buddies into the early hours of the morning made those long nights bearable. Finally, I would like to thank my advisor, Dr. Dunn, for answering every single one of my questions, for directing me through this complex process, for criticizing draft after draft after draft in order to ensure that I did not settle for mediocrity. And most of all, for giving me the good tissues when nothing seemed to be going right. Thank you!

Author's Note

In May 2011, a poster version of this work was presented at the Association for Psychological Sciences, Washington, D.C.

References

Chan, R. C. K., Chen, E. Y. H., Cheung, E. F. C., Chen, R. Y. L., & Cheung, H. K. (2004). Problem-solving ability in chronic schizophrenia: A comparison study of patients with traumatic brain injury. *Eur Arch Psychiatry Clinical Neurosci, 254*, 236-241.

Delis, D., Kaplan, E., & Kramer, J. (2001). *Delis-Kaplan Executive Function System*. San Antonio: The Psychological Corporation, Harcourt Brace & Co.

Doyle, Z., & Dunn, T.M. (2010, May). Response Bias on the Tower Test of the D-KEFS. Poster presented to the Association of Psychological Science, Boston.

Dunn, T. M. (2006). Current issues in response bias during neuropsychological assessment: Incomplete effort to malingering. In: *Focus on Neuropsychology Research* (199-217). Greeley, CO: Nova Science Publishers, Inc.

Leon-Carrion, J., Alarcon, J. C., Revuelta, M., Murillo-Cabezas, F., Domingue-Roldan, J. M., Dominguez-Morales, M.R., Machuca-Murgaa, F., & Forastero, P. (1998). Executive Functioning as outcome in patients after traumatic brain injury. *Intern. J. Neuroscience, 94,* 75-83.

Malik, A., Turner, M. E., & Sadler, C. (2007). Neuropsychological Evaluation. Retrieved October 24, 2009, from *Neuropsychological Evaluation Website* [http://www.medscape.com], *1.* Available: http://emedicine.medscape.com/article/317596-overview

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology,* 22, 213-223.

Tombaugh, T. N. (1996). *TOMM: Test of Memory Malingering.* North Tonawanda, NY: Multi-Health Systems, Inc.