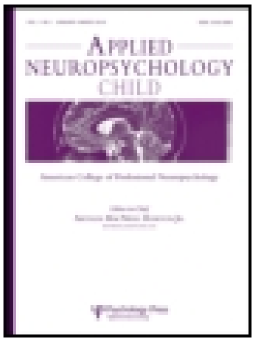

Applied Neuropsychology: Child

ISSN: 2162-2965 (Print) 2162-2973 (Online) Journal homepage: https://www.tandfonline.com/loi/hapc20

Brief communication: Reliability of the D-KEFS Tower Test in samples of children and adolescents in Ecuador

Graham Pluck, Doenya Amraoui & Isabella Fornell-Villalobos

APPLIED NEUROPSYCHOLOGY: CHILD
https://doi.org/10.1080/21622965.2019.1629922

Brief communication: Reliability of the D-KEFS Tower Test in samples of children and adolescents in Ecuador

Graham Pluck, Doenya Amraoui, and Isabella Fornell-Villalobos

Institute of Neurosciences, Universidad San Francisco de Quito, Quito, Ecuador

ABSTRACT

The Tower Test in the Delis-Kaplan Executive Function System (D-KEFS) is a widely-used assessment of executive function in young people. It is similar to other Towers of Hanoi type tasks, for which doubts regarding the reliability of the test have been previously raised. Here, we present data on the internal consistency, unidimensionality, and test–retest reliability of the D-KEFS Towers Test based on an Ecuadorian sample of 264 children and adolescents. In general, the Tower Test appeared to have poor reliability. This may be caused by the combination of very simple and more difficult trials within the same assessment. Nevertheless, one measure, the Time-Per-Move Ratio, appeared to be sufficiently reliable for clinical or research use, and neuropsychological studies have suggested that the Time-Per-Move Ratio may also be the performance measure most sensitive to neurological impairment.

KEYWORDS
Executive function; internal consistency; planning; test reliability; Tower Test

There are various tower tasks used in clinical and research neuropsychology. These are all variations of the Tower of Hanoi puzzle, which originated in mathematics but has come to be a widely-used assessment of cognitive function (Goel, Pullara, & Grafman, 2001). The basic form involves moving disks of different sizes among three points with rules restricting how the discs can be placed. It is usually thought of as a test of frontal lobe function or executive function, particularly planning ability (Shallice, 1982).

One such towers test is provided in the Delis-Kaplan Executive Function System (D-KEFS; Delis, Kaplan, & Kramer, 2001). Apart from supplying rigorous administration guidelines, the D-KEFS also provides normative data from ages 8 to 89. The D-KEFS Tower Test involves nine graded trials from very simple (only one move needed to complete the task) to difficult (a minimum of 26 moves). The main performance measure in each trial is the achievement score. This is based on the number of moves taken to reach the goal state, provided the task is completed within the set time limit. These trial achievement scores are summed to produce the Total Achievement Score. In addition, the D-KEFS manual describes various process scores, such as the number of rule violations across the nine trials (see Table 1 for descriptions of these process scores). These allow clinicians or researchers to focus on particular abilities. For these reasons, the D-KEFS battery, including its Tower Test, has become a popular method of assessment of executive functioning in clinical and research contexts with children and adolescents (e.g., Rose et al., 2017).

We have used the D-KEFS Tower Test in our research on socioeconomic and psychosocial deprivation including vulnerable populations as well as control groups in Quito, Ecuador. However, we have become concerned over the psychometric performance of the Tower Test, particularly the internal consistency and test–retest reliability. Here, we present an analysis of the reliability of the D-KEFS Tower Test based on the data collected in three separate research studies comprising a total of 264 young people.

The publishers of the D-KEFS battery have already provided data on internal consistency of the Tower Test. The metric they chose to use is the split-half reliability coefficient with the Spearman-Brown correction. The split-half coefficients reported in the D-KEFS manual for children and adolescents range between 0.43 and 0.84 for different age groups, with a mean for participants aged 8–15 of 0.61 and for participants aged 16–19 of 0.60. These would be considered ather low, the usual rule-of-thumb being that coefficients over 0.7 are qualitatively “acceptable”, and

CONTACT Graham Pluck gpluck@usfq.edu.ec Institute of Neurosciences, Universidad San Francisco de Quito, Diego de Robles y Vía Interoceánica, Cumbayá, Quito, Ecuador.

Table 1. Descriptions of the various measures of the Tower Test and statistics derived from the test–retest analysis.

Measure	Description	Potential score range	Test mean score (+*SD*)	Retest mean score (+*SD*)	Test–retest correlation
Total Achievement Score	Score based on the number of moves taken. Items 1 and 2 are scored from 0 to 2, items 3 and 4 are scored from 0 to 3, and all later items are scored 0–4. Failure within the time limit scores 0.	0–30	16.68 (2.85)	18.55* (2.69)	$r = .011$
Move Accuracy Ratio	The total number of moves taken across all trials attempted, dived by the minimum number of moves that would be needed for all trials attempted. A perfect performance would therefore by a score of 1, with higher scores indicating less efficient performance.	1–unlimited	1.86 (0.373)	1.81 (0.266)	$r = -.055$
Mean First-Move Time	The mean time between starting a trial and completing a move, for all trials attempted. This indicates how long participants wait before attempting the tasks, low scores may indicate impulsive responding.	0–1,050	2.39 (1.70)	1.97 (0.756)	$r_s = .460$*
Total Rule Violations	The total number of rule violations, i.e. placing a larger disk on a smaller disk, or moving more than one disk simultaneously, across all trials attempted.	0–unlimited	1.32 (1.56)	0.090** (0.426)	N/A
Rule Violations-Per-Item Ratio	The average number of rule violations per trial attempted. High scores indicate frequent rule breaking.	0–unlimited	0.147 (0.173)	0.010** (0.047)	N/A
Time-Per-Move Ratio	The average time taken to make each move, i.e. moving a disk from one location to another.	0–1,050	3.26 (0.810)	2.45*** (0.562)	$r = .756$***

*$p < .05$. **$p < .01$. ***$p < .001$.

over 0.8 are "good." Furthermore, this approach can be criticized because although split-half reliability is a valid measure of internal consistency, different results will be found depending on which items are included in the two halves of the split. Although the test developers used the convention of comparing even numbered items with odd numbered items (as we also did), the fact remains that split-half reliability is limited, as only one of many different possible coefficients is produced. The more common alternative to internal consistency is Cronbach's alpha, which does not suffer from the arbitrariness of the split (Warrens, 2015).

The authors of the D-KEFS manual actually mention that the internal consistency of the Tower Test may be low because of the graded difficulty of the nine trials. Early trials provide training in the correct cognitive set and tap only basic cognitive processes, with later trials tapping higher-level (presumably executive) processes. This may be true. However, if the different trials are measuring different cognitive processes, one could question the logic of summing scores from all completed trials. Summing scores assumes that items are measuring the same construct, i.e. the test is unidimensional. In fact, neither split-half reliability nor Cronbach's alpha provide a good estimate of unidimensionality. An appropriate measure of which would be the mean inter-item correlation and the observation that scores on all or most of the trials are positively correlated (Clark & Watson, 1995). In this brief communication, we examine the internal consistency (split-half and alpha) and the unidimensionality of the Total Achievement Scores on the D-KEFS Tower Test in children and adolescents.

As a measure of reliability, test–retest reliability is usually considered stronger evidence than internal consistency. Executive functions are thought to be stable psychological traits of individuals, as they are highly genetic (Friedman et al., 2008) with most long-term variation being due to development. Therefore, if a test is a good measure of a particular executive function, it should produce similar scores when administered at different time points separated by only a few weeks. The developers of the D-KEFS Tower Test provide a test–retest reliability coefficient on a sample of 25 individuals aged from 8 to 19 for the primary performance measure, the Total Achievement Score. Their estimate of the test–retest reliability is $r = .51$, which again is lower than would usually be desirable in a neuropsychological test for clinical or research use. However, they do not provide test–retest coefficients for any of the five process scores. In this research, we provide some of these estimates based on a subsample of 22 adolescents who were assessed twice with the D-KEFS Tower Test with a delay of about five weeks between sessions.

Method

The participants in this analysis were recruited for three separate neuropsychological studies. Two of these are currently unpublished. One of which is a study of children living in foster care and a control group of school children with no histories of foster care. The children in foster care were all living in one of five different group homes in the city of Quito, Ecuador, because of removal from their families due to abuse/neglect or because they were orphaned. The control sample for the foster care study was recruited from two different state-run schools, and all participants were screened for past experiences of foster care, child labor or homelessness. The other unpublished work is a study of socioeconomic status (SES) and neuropsychological functions of children. None of the participants were selected for any specific vulnerability and represent children from a range of SES backgrounds. These were recruited mainly from 10 different schools around the city of Quito, Ecuador. However, one of these schools only catered for boys, contributing to an overrepresentation of male participants in this sample. The third study was a neuropsychological study of "street children" and a control sample of school children which has already been published (Pluck, Banda-Cruz, Andrade-Guimaraes, & Trueba, 2018). The participants were in fact all former street children attending a non-governmental educational program. They were all living with their families at the time of recruitment. The control sample was recruited from a single state-run school in Quito, Ecuador. The data from that study is reanalyzed here to extend the size of the overall sample. Demographic details of the different samples are shown in Table 2. The table also includes details of "All special" (participants who were recruited because of special vulnerabilities i.e., street children or in foster care), "All controls" (i.e., all children who were recruited in the studies without being selected for any special vulnerability such as being street children or in foster care), and "Combined sample" (i.e., all participants).

Table 2. Details of the different samples and subsamples analyzed and the Cronbach's α values for the Tower Test Total Achievement Score.

	Total *n*	Female *n* (%)	Mean age	Age range	α
Street children	36	9 (25%)	13.48	10.71–16.16	0.623
Foster care	43	18 (42%)	15.93	13.29–18.39	0.125
All special	79	27 (34%)	14.81	10.71–18.39	0.540
Controls for street children	24	11 (46%)	13.95	10.33–16.50	0.354
Controls for foster care	48	25 (52%)	16.23	12.90–21.72	0.491
SES sample	113	36 (32%)	14.94	11.85–19.01	0.605
All controls	185	73 (39%)	15.15	10.33–21.72	0.567
Combined sample	264	99 (40%)	15.05	10.33–21.72	0.576

Note. All special is a group composed of all of the street children and all of the foster care participants. The SES sample is composed of participants recruited to a study of socioeconomic status (not selected for any specific vulnerabilities). All controls is composed of the control groups for the street children and foster care studies, as well as the SES sample. The Combined sample is composed of all the participants from the different studies.

All studies had institutional research ethics committee approval. Written informed consent was provided by parents or guardians of all participants, and written assent was provided by all of the participants. They were then assessed with the D-KEFS Tower Test following the standard instructions. The primary performance measure (Total Achievement Score) and the individual trial achievement scores were recorded for all participants. However, different secondary process scores were recorded in the different studies. Therefore, the data on these is not available for the internal consistency calculations.

Twenty-two participants in the third study (SES) were tested twice. The mean delay between test sessions was 33.8 days (range 21–54). For these individuals, all of the primary performance and secondary processes measures were recorded. The mean age of this sample was 16.39 years, range 12.01–19.01, 8/22 female.

Results and discussion

For the Total Achievement Score the internal consistency as estimated by the split-half method (odd-numbered verses even-numbered items) with Spearman-Brown correction (for unequal lengths) was found to be 0.577. This is similar to the values reported in the D-KEFS manual of around 0.600 for the same calculation on their normative sample. We also calculated the Cronbach's alpha based on standardized items (not provided in the D-KEFS manual); this produced a very similar estimate to that of our split-half analysis, with an alpha of 0.576 for the combined sample of 264 participants. We also examined the alpha values for the different subgroups, which are shown in Table 2. The different estimates of internal consistency ranged from 0.125 to 0.623 across the samples. This includes a value of 0.567 for all of the control participants.

Although the Cronbach's alpha value in the combined sample is 0.576, the D-KEFS Tower Test includes a stop rule (after three failures). This potentially produces dependence between scores on earlier and later items that will inflate estimates of internal consistency. If we exclude the 18 cases in which the stop rule was used, the Cronbach's alpha drops from 0.576 to 0.369, and the split-half reliability drops from 0.577 to 0.300.

As previously mentioned, this low internal consistency as measured by Cronbach's alpha could be due

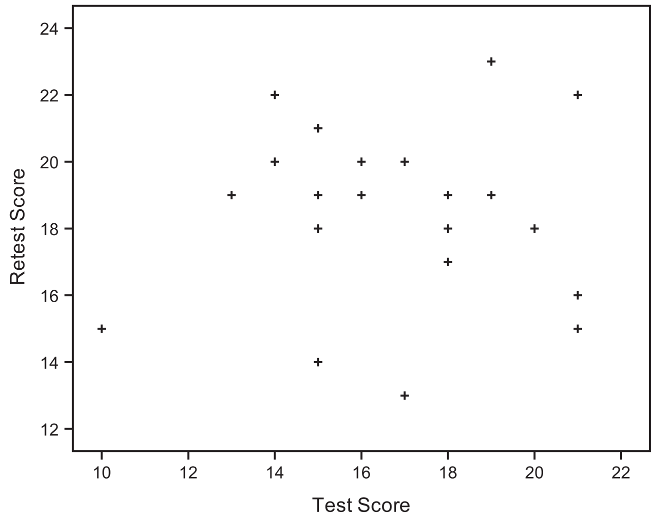

Figure 1. Scatterplot showing the relationship between Total Achievement Scores on the D-KEFS Tower Test at test and retest.

to the combination of simple and complex tasks. However, we found that sequentially removing the simpler items and recalculating does not increase the Cronbach's alpha value. In fact, removing the easiest items one at a time until only the three most difficult tasks remained produced a further reduction in the alpha value at every step. Furthermore, the combination of simple and complex tasks cannot explain the low internal consistency as measured by split-half reliability.

We also explored whether the wide age range of our participants could be affecting the internal consistency. We did this by excluding outliers and producing two age-range groups: the group of 11–14 year-olds ($n = 132$) was found to have an alpha of 0.627 and the group of 15–18 year-olds ($n = 125$) was found to have an alpha of 0.514. This suggests that internal consistency may be better in the younger age groups. However, if we again exclude cases in which the stop rule was used the alpha values again drop to 0.366 for the younger and to 0.377 for the older participants.

We also investigated the unidimensionality of the Total Achievement Score. This is a feature of a test that cannot be inferred directly from the internal consistency. The mean inter-item correlation for the full sample for the Total Achievement Score was found to be 0.131 (range of r values = −.032 to +.241). This would usually be considered too low and too variable for the scale to be considered unidimensional, in which 0.15 is considered the lower limit (Clark & Watson, 1995).

Regarding the test–retest reliability, we investigated scores on not only the primary measure (Total Achievement Score) but also on the five process scores. This was performed with simple bivariate correlations, and the results are shown in Table 1. For the Total Achievement Scores, there was no apparent association between scores on the first and second administrations. To highlight this, the scatterplot is shown in Figure 1. This seems to show very poor temporal stability of the primary performance measure on the D-KEFS Tower Test in young people. Previous researchers have also reported low test-retest reliabilities of tests supposed to tap executive functions, and it has been suggested that tests in which there may be a realization that improves performance are particularly susceptible to this (Lowe & Rabbitt, 1998). This would be particularly pertinent to Tower of Hanoi type tests, as various rules can be learnt which guarantee success if followed (Simon, 1975). Nevertheless, Piovesana, Ross, Whittingham, Ware, and Boyd (2015) have reported an acceptable level of test–retest reliability of the D-KEFS Tower Test in children with cerebral palsy (r = .74). In addition, a study with a Towers of Hanoi task in 45 healthy children aged 7–10 found a test–retest correlation of 0.500 (Bishop, Aamodt-Leeper, Creswell, McGurk, & Skuse, 2001). This is similar to the value of 0.510 reported in the D-KEFS manual for a group of 25 8–19 year-olds. Again, these scores are less than desirable, but they are still much higher than found in the current report. There are several factors that may partly explain the discrepancy between test–retest reliability estimates in the D-KEFS normalization sample and the current sample for the Tower Test Total Achievement Scores.

One possibility is that the test-retest delay is important. For the results reported in the D-KEFS manual, the mean delay was 25 days, while in the

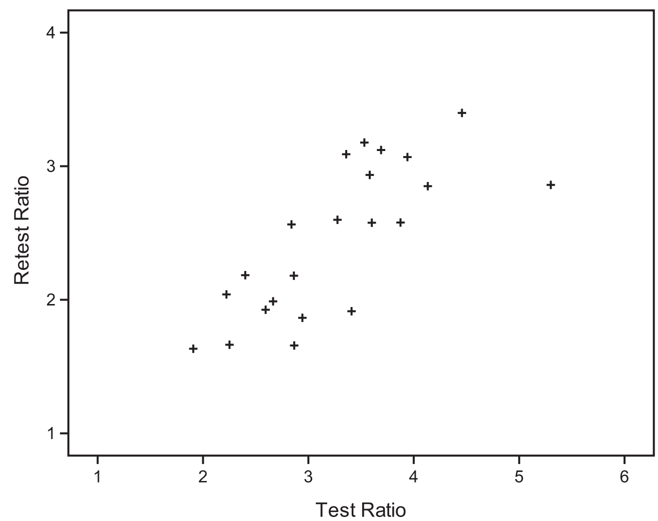

Figure 2. Scatterplot showing the relationship between Time-Per-Move Ratios for the D-KEFS Tower Test at test and retest.

current report it was 34 days. Differences in the compositions of the two samples may also be a key factor. The age range in the D-KEFS sample was 11 years, whereas in the current report it was only 7 years. It is also possible that the test is interpreted differently across cultures. The D-KEFS manual reports on a USA-based sample, while in the current research the participants were all Ecuadorian. Indeed, research with adults has shown that there are important cross-cultural differences on performance of the D-KEFS battery (Kelkar, Hough, & Fang, 2013).

It should also be noted that the D-KEFS Tower Test is much shorter than many other Tower of Hanoi type tasks. Keifer and Tranel (2013) have suggested that the precision of the D-KEFS Tower Test is limited by an insufficient number of medium difficulty trials. This could produce temporal instability as well as limit the internal consistency. However, it cannot explain why our test–retest reliability estimate is much lower than that reported in the D-KEFS manual, as the same test was used in both studies. It is also possible that D-KEFS Tower Test has less reliability in normal samples, such as reported here, compared to clinical samples, such as reported by Piovesana et al. (2015). This is because data sets from normal samples typically contain less variation than data from clinical samples (e.g., in our sample the SDs for the Total Achievement Score are <3, and for Piovesana et al., 2015, they are >4). In samples with lower variance, correlation values are usually suppressed compared to samples with greater variance (Howell, 1992).

Regarding the process scores, there was again no apparent association between test and retest performance for the Move Accuracy Ratio scores. This is perhaps not surprising as it is an accuracy derived measure and may have similar properties to the Total Achievement Score previously discussed. In contrast, there is a medium-sized positive correlation between test and retest Mean First-Move Time scores. Nevertheless, at $r_s = .460$ this is still somewhat below desirable levels for reliability. We had intended to analyze two other process measures of the D-KEFS Tower Test, Total Rule Violations and Rule Violations-Per-Item Ratios. However, although the majority of participants (13/22) made at least one rule violation in the test phase, only two participants made a single rule violation each in the retest phase. This precludes correlational analysis, but it does suggest that test and retest performances are quite different with regard to rule violations. Finally, there was a large positive correlation between test and retest performance for the Time-Per-Move Ratio scores (see Table 1 and Figure 2). Therefore, this appears to be the only measure on the D-KEFS Tower Test with acceptable temporal stability in terms of correlation. This result is in accordance with previous findings of test–retest reliability of neuropsychological tests, which suggest that accuracy-based measures of performance are generally less reliable than time-based measures (Lowe & Rabbitt, 1998).

Although correlation values give a good estimate of relative stability over time, they are uninformative about absolute changes in performance (e.g., practice effects). From Table 1, it seems that performance tended to improve from the test to retest sessions. This was confirmed with paired sample *t*-tests. For the Total Achievement Score there was a significant improvement over test sessions, $t(21) = -2.251$, $p = .035$, $d = 0.480$. This indicates a practice effect of similar magnitude to the practice effect previously reported with the Tower Test in a group of children with cerebral palsy (Piovesana et al., 2015). There was also a significant practice effect in the current study for the Time-Per-Move Ratio, $t(21) = 7.143$, $p < .001$, $d = 1.523$. This was in fact a qualitatively large practice effect; the usual threshold for a 'large' effect is $d = 0.8$, which was well surpassed in the current analysis. However, there were no significant differences between test and retest performances for the Move Accuracy Ratio and the Mean First-Move Time process measures, ($ps > .255$). In contrast, nonparametric paired-sample analyses confirmed what was previously noted, both Total Rule Violations and Rule-Violations-Per-Item were significantly less frequent in the retest session compared to the test session (both $p = .002$).

Thus, significant practice effects are present on several measures of the Tower Test, including the measure with the best test-retest reliability based on the correlation analysis, the Time-Per-Move Ratio. This does not diminish the reliability of that particular measure for single assessments, but suggests that repeated assessments may be prone to substantial change due to practice.

Although we found only one measure which appears to have acceptable test-retest reliability based on the correlations, the Time-Per-Move Ratio, high scores on this measure may actually be one of the most valid measures of neuropsychological impairment on the Tower Test. This is because in one study comparing adult patients with lateral prefrontal lesions against a control sample for all D-KEFS Tower Test measures, the largest effect size between case and control performance was for the Time-Per-Move Ratio (Yochim, Baldo, Kane, & Delis, 2009). Similarly, a study of children with congenital heart disease, which has numerous deleterious effects of brain development, found that the Time-Per-Move Ratio again had the highest effect size for distinguishing cases from controls (Cassidy, White, DeMaso, Newburger, & Bellinger, 2015). Overall, when using the D-KEFS Tower Test, the Time-Per-Move Ratio measure may be the most reliable and valid measure of neuropsychological status, with higher scores indicative of impairment. However, even that measure is subject to substantial practice effects, which may limit its use for monitoring change over time in clinical practice.

Nevertheless, some limitations of the current research should be acknowledged. The samples analyzed were not selected specifically for testing the reliability of the Tower Test and do not constitute a normative sample. In particular, two of the samples reported were recruited because of psychosocial vulnerabilities, that is, being former street children or living in foster care. Conversely, we report on a large overall sample, and the subgroup analyses reveal a fairly consistent pattern of low internal consistency regardless of subgroup (e.g., street child, foster, or control). Considering the large number of published research studies of executive function in children, there are remarkably few studies of the psychometric properties of common assessments such as the Tower Test. The current research partly remedies this imbalance.

Overall, the current results suggest caution when using measures other than the Time-Per-Move Ratio with the D-KEFS Tower Test for clinical or research purposes in young people. In particular, as most of the measures have low test–retest reliability and are subject to practice effects, repeat testing with the D-KEFS Tower Test over time may give misleading impressions of clinical change.

Acknowledgments

We would like to thank all of the young people who participated in our research. The studies described in this article did not receive any funding.

Disclosure statement

The authors declare no conflict of interest.

References

Bishop, D. V., Aamodt-Leeper, G., Creswell, C., McGurk, R., & Skuse, D. H. (2001). Individual differences in cognitive planning on the Tower of Hanoi task: Neuropsychological maturity or measurement error? *Journal of Child Psychology and Psychiatry*, *42*(4), 551–556. doi:10.1111/1469-7610.00749

Cassidy, A. R., White, M. T., DeMaso, D. R., Newburger, J. W., & Bellinger, D. C. (2015). Executive function in children and adolescents with critical cyanotic congenital heart disease. *Journal of the International Neuropsychological Society*, *21*(1), 34–49. doi:10.1017/S1355617714001027

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319. doi:10.1037/1040-3590.7.3.309

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan executive function system: Technical manual*. San Antonio, TX: Psychological Corporation.

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

Goel, V., Pullara, D., & Grafman, J. (2001). A computational model of frontal lobe dysfunction: Working memory and the Tower of Hanoi task. *Cognitive Science*, *25*(2), 287–313. doi:10.1207/s15516709cog2502_4

Howell, D. C. (1992). *Statistical methods for psychology*. Belmont, CA: Duxbury Press.

Keifer, E., & Tranel, D. (2013). A neuropsychological investigation of the Delis-Kaplan Executive Function System. *Journal of Clinical and Experimental Neuropsychology*, *35*(10), 1048–1059. doi:10.1080/13803395.2013.854319

Kelkar, A. S., Hough, M. S., & Fang, X. (2013). Do we think alike? A cross-cultural study of executive functioning. *Culture and Brain*, *1*(2–4), 118–137. doi:10.1007/s40167-013-0010-4

Lowe, C., & Rabbitt, P. (1998). Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. Cambridge Neuropsychological Test Automated Battery. International Study of Post-Operative Cognitive Dysfunction. *Neuropsychologia*, *36*(9), 915–923. doi:10.1016/S0028-3932(98)00036-0

Piovesana, A. M., Ross, S., Whittingham, K., Ware, R. S., & Boyd, R. N. (2015). Stability of executive functioning measures in 8-17-year-old children with unilateral cerebral palsy. *The Clinical Neuropsychologist*, *29*(1), 133–149. doi:10.1080/13854046.2014.999125

Pluck, G., Banda-Cruz, D. R., Andrade-Guimaraes, M. V., & Trueba, A. F. (2018). Socioeconomic deprivation and the development of neuropsychological functions: A study with "street children" in Ecuador. *Child Neuropsychology*, *24*(4), 510–523. doi:10.1080/09297049.2017.1294150

Rose, M., Reville, M. C., Iszatt, A., Levinson, S., Frampton, I., & Lask, B. (2017). Deconstructing planning ability in children and adolescents with anorexia nervosa. *Applied Neuropsychology: Child*, *6*(4), 297–304. doi:10.1080/21622965.2016.1172312

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *298*(1089), 199–209. doi:10.1098/rstb.1982.0082

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, *7*(2), 268–288. doi:10.1016/0010-0285(75)90012-2

Warrens, M. J. (2015). *On Cronbach's alpha as the mean of all split-half reliabilities Quantitative Psychology Research – The 78th Annual Meeting of the Psychometric Society*. New York: Springer.

Yochim, B. P., Baldo, J. V., Kane, K. D., & Delis, D. C. (2009). D-KEFS Tower Test performance in patients with lateral prefrontal cortex lesions: The importance of error monitoring. *Journal of Clinical and Experimental Neuropsychology*, *31*(6), 658–663. doi:10.1080/13803390802448669