

---

# Heritability of Wisconsin Card Sorting Test (WCST) and Stroop Color-Word Test Performance in Normal Individuals: Implications for the Search for Endophenotypes

Jeanette Taylor

Department of Psychology, Florida State University, Tallahassee, Florida, United States of America

A surge in the search for endophenotypes for psychiatric disorders has occurred in the past several years. An important criterion of an endophenotype is that it is heritable. Two of the most widely used executive cognitive functioning measures are the Wisconsin Card Sorting Test (WCST) and the Stroop Color-Word Test. Each has been considered as a possible endophenotype. However, research on the heritability of each of these measures is sparse, and in the case of the WCST, mixed. As part of a pilot twin study examining cognitive functioning and personality in adults, the WCST and the Stroop were administered to 80 monozygotic and 29 dizygotic twin pairs screened for absence of neurological disease and head injury. Results replicated and extended previous findings for moderate heritability of Stroop performance. However, the WCST showed little evidence of genetic influence, suggesting that it might not meet one of the criteria for an endophenotype.

---

Poor executive cognitive functioning has been associated with a variety of psychiatric disorders, including schizophrenia (see Szoke et al., 2005 for a recent meta-analysis), co-morbid substance dependence and conduct disorder (Giancola & Mezzich, 2000), and attention-deficit/hyperactivity disorder (ADHD; Doyle et al., 2005). Executive functioning is a broad domain encompassing many dimensions, including response inhibition, focusing attention on relevant stimuli while tuning out extraneous ones, organizing and planning, updating, and decision-making (Miyake et al., 2000; Stuss & Alexander, 2000). The apparent associations with a specific area of the brain (Fishbein, 2000), coupled with their association with psychiatric disorders, make certain dimensions of executive functioning obvious candidates for endophenotypes or 'biological intermediate traits' between genes and psychiatric phenotypes (Almasy & Blangero, 2001; de Geus, 2002; Gottesman & Gould, 2003; Inoue & Lupski, 2003).

Two of the most widely used executive functioning measures are the Wisconsin Card Sorting Test (WCST), which taps cognitive flexibility and ability to shift set, and the Stroop Color-Word Test, which taps cognitive inhibition of an overlearned response. Performance on each task has been linked to different areas of the frontal cortex (Fishbein, 2000), which make them each potentially valuable candidates for endophenotypes. Not surprisingly, each task has been considered as a possible endophenotype for disorders such as schizophrenia (Keri & Janka, 2004) and ADHD (Doyle et al., 2005; Stins et al., 2004).

The identification of endophenotypes can enhance the search for genes associated with behavioral phenotypes because the individual genes in a polygenic system would account for more variance in an endophenotype than in the behavioral phenotype (Almasy & Blangero, 2001; de Geus, 2002; Gottesman & Gould, 2003; Inoue & Lupski, 2003), though recent work suggests that endophenotypes might have complex genetic architectures much like the phenotypes they are meant to replace (Flint & Munafò, 2007). Performance on the WCST has shown mixed findings with regard to associations with specific genes. For example, it has been associated with Catechol-O-Methyltransferase (COMT) in healthy (e.g., Bruder et al., 2005) and schizophrenic subjects (e.g., Egan et al., 2001; Rybakowski, Borkowska, Czerski et al., 2006) and with BDNF gene polymorphisms in bipolar (but not schizophrenic) subjects (Rybakowski, Borkowska, Skibinska et al., 2006). However, two studies have failed to find an association between WCST and COMT polymorphisms in children with ADHD (Mills et al., 2004; Taerk et al., 2004). Stroop performance has been associated with interactions between DRD2 and COMT genes (Reuter et al., 2005). However, these genetic association findings do

---

Received 24 August, 2007; accepted 13 September, 2007.

Address for correspondence: Dr Jeanette Taylor, Florida State University, Department of Psychology, PO Box 3061270, Tallahassee, FL 32306-1270. E-mail: taylor@psy.fsu.edu.

not in themselves provide evidence that WCST or Stroop performance is suitable as an endophenotype. An endophenotype has several characteristics (Almasy & Blangero, 2001; de Geus, 2002; Gottesman & Gould, 2003; Inoue & Lupski, 2003), including (1) it is correlated with the disease and/or disease severity, (2) it is heritable, (3) it is found in unaffected family members, (4) its association with the disease is state-independent, and (5) it is reliable and stable. Doyle et al. (2005) outline the evidence for executive functioning measures as potential endophenotypes for ADHD, and note that information on heritability of many of the measures is lacking or mixed. This is certainly the case for the widely used WCST and Stroop executive functioning measures.

Only four published studies have examined the heritability of WCST performance. Two studies were based on small samples (30 or fewer twin pairs) and found no evidence of genetic influence on WCST scores (Campana et al., 1996, Nicole & Del Miglio, 1997). Using a somewhat larger sample (58 MZ, 25 DZ) of female young adult twins, Anokhin et al. (2003) showed moderate heritability for most (not all) WCST scores. Finally, like the earlier studies, Kremen et al. (2007) found no evidence of genetic influence on any WCST score in a large subsample (over 300 twin pairs) of middle aged male twins drawn from the Vietnam Era Twin Registry. Thus, the literature on the heritability of WCST performance is small and mixed, with the two largest twin studies showing opposite results from quite different samples (young adult women vs. middle aged men).

The literature on the heritability of Stroop performance is similarly small, though the results are fairly consistent. Johnson et al. (2003) examined Stroop performance in 50 pairs of MZ and 37 pairs of DZ twins reared apart (the majority of the sample was age 40–70). The MZ twins reared apart correlation is a direct estimate of heritability. It ranged from .43 to .53 for the three Stroop trials (word reading, color reading, and color-of-the-word reading), and it was .34 for the interference score calculated from the three trials. Stins et al. (2004) examined response times on the Stroop test in 290 12-year-old twins, and reported a heritability of about 70% for the three Stroop trial scores, and 50% for the interference score. In sum, the literature on the heritability of WCST performance is small and mixed, and the two published studies examining heritability of Stroop performance are consistent, but cover nonoverlapping age groups.

The present study examined both WCST and Stroop performance in male and female adult twins, to help clarify research on the heritability of the former, and round out evidence for heritability in the latter, within an age group missed by previous studies.

## Materials and Methods

### Twin Ascertainment

Twins in this study were recruited during the creation of the Florida State Twin Registry (Taylor et al., 2006). Most twins in the registry were ascertained

through the registrar at Florida State University, based on a match on last name, date of birth, and city of birth provided to the investigator each semester. Potential twin pairs identified through the registrar matches were contacted via mail to confirm that they were twins and, if so, to invite participation in the study and in the registry. Twins from the university and from the surrounding communities also volunteered for the study after learning about it from advertisements or word of mouth.

### Study Sample

The sample for the study included 80 MZ pairs (22 male) and 29 same-sex DZ pairs (4 male). Twins in the study ranged in age from 18 to 83 years old ( $M = 23.76$ ;  $SD = 11.68$ ). The MZ pairs were not significantly different in age ( $M = 23.73$ ;  $SD = 12.27$ ; range in age: 18–83) than the DZ pairs ( $M = 23.85$ ;  $SD = 10.18$ ; range in age: 18–60). The sample racial/ethnic composition reflected that of the University and the surrounding area: 77% white; 16% African-American; 5% Hispanic; and 3% mixed or other race.

Zygosity was determined using two methods: a 5-item self-report questionnaire, completed by the twins, that has over a 95% accuracy rate as indicated by validity testing using DNA (Lykken et al., 1990), and a rating of twin similarity adopted from the one used on the Minnesota Twin Family Study on natural hair color, natural eye color, and ear shape and attachment, completed by trained research staff. At the time of the similarity rating, a close-up digital photograph was taken of an eye and an ear of each twin — this was used in resolving any discrepancies between the similarity rating and the self-reported zygosity questionnaire. Discrepancies in twin and staff ratings occurred in 8% of the sample. Most discrepancies could be resolved by examining the digital photographs; those that could not be resolved were coded based on the twin reported zygosity questionnaire.

### Study Procedure

Twins completed the study at the same time during a single session in a laboratory on the Florida State University campus. All study procedures were approved by the IRB of Florida State University, and participants were treated in accordance with the standards of the granting agency that funded the research. Written informed consent was obtained from each twin prior to beginning any procedures. A research assistant conducted the twin similarity rating, and then the twins were separated to independently complete the remaining procedures. Cognitive functioning measures were administered first, followed by self-reports and interviews. The study took approximately 3.25 hours to complete, and each twin received US\$30.00 for participating. Only the cognitive functioning measures were examined for this report.

### Study Measures

Participants completed the WCST-64: Computer Version for Windows — Research Edition (Heaton, 2000). The

WCST requires the respondent to sort cards according to a predetermined criterion (color, form, or number) using only minimal feedback ('Correct' or 'Incorrect') after each trial. The computer program yields *T* scores ( $M = 50$ ;  $SD = 10$ ) for Total Errors, Perseverative Responses, Perseverative Errors, and Nonperseverative Errors. These *T*-scores are normed for age (and, for those over 20, age and education level; Heaton et al., 1993). Raw scores for the number of categories completed are also provided. Participants were also administered the Stroop Color-Word Test (Golden, 1978) by a trained research assistant. The Stroop test begins with a word reading test (Word) followed by a color naming test (Color) and finally a test that requires the respondent to name the color that the word is printed in, ignoring the content of the word (Color-Word). On each card, the participant is given 45 seconds to read the items out loud as quickly as possible. If a mistake is made, the test administrator indicates this, and the participant must correct the error before continuing. The raw score for each card is the number of items read within the allotted 45 seconds. The Stroop yields a *T* score for each test, as well as an Interference score calculated from the other test scores (Golden, 1978).

#### Data Analysis

For each variable, twin intraclass correlations, means, and standard deviations were calculated. Then, univariate biometrical models were fit to the data to estimate additive genetic (A), shared environmental (C), and non-shared environmental (E) influences. (Given the small sample size, models testing nonadditive genetic effects were not fit to the data, as the sample was certainly underpowered to detect such effects.) Models were fit to variance-covariance matrices using maximum likelihood estimation in Mx (Neale et al., 1998). Expected covariances were specified as follows:

$$MZ \text{ cov} = A + C$$

$$DZ \text{ cov} = .5A + C$$

Overall model fit was evaluated with the goodness-of-fit chi-square statistic. The results from the full (ACE) model informed decisions about which nested models to fit to the data. Nested models (e.g., AE) were compared to the full model by testing the significance of the difference in chi-square, which is the difference in  $-2 \log$ -likelihood, between the two models on the difference in *df* between models. A nonsignificant chi-square difference indicated that the nested model could be selected over the full model. Competing nested models were compared to each other using two fit indices, Akaike's criterion (AIC; Akaike, 1987) and Root Mean Squared Error Approximation (RMSEA), which balance model fit with parsimony. Lower values of AIC indicate better fit, and RMSEA less than .05 indicates excellent model fit (values of .06–.10 indicate adequate fit). Thus, the nested model that minimized both AIC and RMSEA was selected as the best-fitting model.

#### Results

Data were missing on the Stroop for one pair of male DZ twins, due to color blindness in one of the twins, yielding a total of 28 pairs available for analyses. There was a significant correlation between age and WCST and Stroop scores. Gender also correlated significantly with Stroop scores. As such, data on each measure were age- and sex-corrected as needed, prior to analyses using regression procedures outlined by McGue and Bouchard (1984). Given the potential concerns about using age- and sex-corrected *T* scores, analyses were repeated using age- and sex-corrected raw scores, and the results matched those obtained with *T* scores. Results were also the same when treating raw variables as ordinal after creating six

**Table 1**

Raw and *T* score Means (and Standard Deviations) and Intraclass Correlations by Zygosity

	MZ Raw Mean (SD)	DZ Raw Mean (SD)	MZ <i>T</i> -score Mean (SD)	DZ <i>T</i> -score Mean (SD)	rMZ (95% CI)	rDZ (95% CI)
WCST	( <i>n</i> = 160)	( <i>n</i> = 58)	( <i>n</i> = 160)	( <i>n</i> = 58)	( <i>n</i> = 80 pairs)	( <i>n</i> = 29 pairs)
Number of Categories						
Completed	3.50 (1.44)	3.93 (1.35)	—	—	.01 (–.21, .23)	–.26 (–.57, .11)
Total Errors	15.81 (9.37)	13.88 (8.96)	50.54 (10.99)	52.00 (10.04)	.17 (–.05, .37)	–.36 (–.64, .01)
Perseverative Responses	8.35 (6.38)	7.60 (5.23)	50.51 (10.73)	50.04 (9.06)	.04 (–.18, .26)	.04 (–.40, .33)
Perseverative Errors	7.74 (5.08)	7.05 (4.52)	50.29 (11.11)	50.24 (9.32)	.07 (–.15, .28)	.00 (–.37, .37)
Nonperseverative Errors	8.07 (6.04)	6.83 (5.00)	49.43 (10.47)	51.09 (9.64)	.21 (–.01, .41)	–.49 (–.73, –.15)
Stroop Color-Word Test	( <i>n</i> = 160)	( <i>n</i> = 56)	( <i>n</i> = 160)	( <i>n</i> = 56)	( <i>n</i> = 80 pairs)	( <i>n</i> = 28 pairs)
Word	105.56 (15.94)	107.02 (13.67)	48.81 (7.95)	49.48 (6.86)	.54 (.36, .68)**	.01 (–.36, .38)
Color	81.48 (12.09)	79.38 (9.51)	51.05 (8.04)	49.59 (6.22)	.58 (.41, .71)*	–.05 (–.41, .33)
Color-Word	50.41 (9.46)	49.55 (8.43)	55.21 (9.52)	54.16 (9.42)	.59 (.42, .72)**	.01 (–.36, .38)
Interference	4.62 (7.72)	4.14 (6.83)	—	—	.39 (.18, .56)	.18 (–.20, .52)

Note: \* $p < .05$  MZ vs. DZ difference; \*\* $p < .01$  MZ vs. DZ difference. WCST = Wisconsin Card Sorting Test.

**Table 2**

Univariate Biometrical Model Fitting Results for the Stroop Color-Word Test

	A	C	E	$\chi^2$ ( <i>df</i> )	<i>p</i>	AIC	RMSEA	$\chi^2 \Delta$ ( <i>df</i> )
Word Score	.50 (.00–.67)	.00 (.00–.46)	.50 (.36–.67)	2.904 (3)	.407	–3.096	.077	—
	<b>.50 (.33–.64)</b>	—	<b>.50 (.36–.67)</b>	<b>2.904 (4)</b>	<b>.574</b>	<b>–5.096</b>	<b>.046</b>	<b>0.72 (1)</b>
Color Score	.53 (.00–.66)	.00 (.00–.51)	.47 (.34–.63)	9.275 (3)	.026	3.275	.199	—
	<b>.53 (.37–.66)</b>	—	<b>.47 (.34–.63)</b>	<b>9.275 (4)</b>	<b>.055</b>	<b>1.275</b>	<b>.166</b>	<b>0.71 (1)</b>
Color-Word Score	.57 (.19–.69)	.00 (.00–.34)	.43 (.30–.59)	10.843 (3)	.013	4.843	.239	—
	.57 (.41–.69)	—	.43 (.30–.59)	10.843 (4)	.028	2.843	.201	—
Interference	.17 (.00–.53)	.20 (.00–.51)	.63 (.46–.81)	5.056 (3)	.168	–0.944	.138	—
	<b>.38 (.19–.53)</b>	—	<b>.62 (.47–.81)</b>	<b>5.250 (4)</b>	<b>.263</b>	<b>–2.750</b>	<b>.113</b>	<b>0.19 (1)</b>
	—	.36 (.18–.51)	.64 (.48–.82)	5.205 (4)	.267	–2.795	.113	0.15 (1)

Note: Standardized parameter estimates (and 95% CIs) are provided for each model. For each variable, the ACE model results are presented first followed by the AE model results. For the Interference score an additional model (CE) was fit and is presented last. The best-fitting model is indicated in bold type. Neither of the models fit for the Color-Word Score.

categories for each. Moreover, the need for additional age correction of the scores could raise potential concerns about the inclusion of older adult twins in this mostly young adult sample. As such, the analyses were repeated (including a new round of age- and sex-correction as needed) after removing all twins over the age of 40, and the results closely matched those obtained from the full sample. As such, the results from the full sample using age- and sex-corrected *T* scores are presented.

The distribution for each variable was normal and none required transformation. For each variable, all but a few twins had scores within 2.5 *SD* of the mean. (Note that outlier scores were adjusted by bringing them to within 2.5 *SD* of the mean, and results closely matched those from analyses in which scores were not adjusted.) Table 1 presents (noncorrected) raw and *T* score means and correlations for each variable by zygosity. No significant mean or variance differences between MZ and DZ twins were found for any variable. Consistent with other research (e.g., Miyake et al., 2000), performance on the Stroop and WCST was at best modestly correlated (all *r*s less than .16 absolute value) in the entire sample. Twin correlations in Table 1 suggested that WCST performance is not heritable. As such, biometrical models were fit only to the Stroop scores and those results are presented in Table 2.

The AE nested model provided the best fit for the Word and Color scores. The models did not adequately fit the data for the Color-Word score, though the pattern and magnitude of the estimates in the models were consistent with those from previous studies. Finally, the Interference score showed possible shared environmental influence, but the AE and CE nested models were essentially the same, and the AE model was more consistent with the twin correlations.

## Discussion

The recent surge in research on endophenotypes for certain psychiatric disorders (e.g., schizophrenia) has

focused in part on executive functioning. Two widely used measures of executive functioning, the WCST and the Stroop test, have each received attention as possible endophenotypes (Doyle et al., 2005; Keri & Janka, 2004; Stins et al., 2004). However, the evidence for genetic influence on those measures is limited and/or mixed, which has implications for the use of either measure as an endophenotype, given that heritability is a *necessary* characteristic of an endophenotype (Almasy & Blangero, 2001; de Geus, 2002; Gottesman & Gould, 2003; Inoue & Lupski, 2003). The present study aimed to clarify and round out evidence on the heritability of these two widely used executive functioning measures.

No evidence of genetic influence on WCST performance was found. Consistent with an earlier twin study by Campana et al., 1996, the correlations on most WCST scores for MZ twins were near zero and non-significant, and the DZ correlations were zero or modestly negative and nonsignificant. The two earlier twin studies on the WCST (Campana et al., 1996, Nicole & Del Miglio, 1997) were based on very modest samples, and were contrasted in their results by a study of female young adult twins that did find evidence of moderate genetic influence on most WCST scores (Anokhin et al., 2003). However, the present findings from a larger mixed gender sample of mostly young adult twins, coupled with similarly negative findings from a larger sample of middle-aged male twins (Kremen et al., 2007), appear to tip the balance of evidence toward suggesting that WCST performance is not heritable. This is not necessarily incompatible with findings of specific genetic associations with WCST performance. Indeed, the COMT gene polymorphism accounts for less than 5% of the variance in WCST performance (Egan et al., 2001) and, therefore, it is possible that twin studies examining WCST performance have been underpowered to detect what might be very modest magnitude heritability for the task.

Stroop performance, on the other hand, had clear genetic influence in this study of mostly young adult twins, with estimates of additive genetic effects coming

close to those found in preadolescent twins (Stins et al., 2004) and middle- to late-adult twins (Johnson et al., 2003). As such, the present data round out the literature on the heritability of the Stroop test, and suggest that the magnitude of genetic effects are fairly consistent from late childhood to late adulthood (though this remains an empirical question).

So what are the implications of the present results on the search for endophenotypes? The Stroop test appears to meet the heritability criterion and, as Stins et al. (2004) suggest, might represent an endophenotype for dysfunction in the interference dimension of executive functioning. The WCST, on the other hand, does not appear heritable based on the balance of the evidence, and thus fails to meet a critical criterion for an endophenotype. This certainly does not diminish the impact of the research showing impaired functioning on the WCST in relation to schizophrenia, ADHD, or other psychiatric disorders. Instead, it suggests simply that normal variation in WCST performance is not associated with large variation in genetic factors, and thus other explanations for its association to psychiatric phenotypes require investigation. Gottesman and Gould (2003) make a distinction between a 'biological marker' which lacks genetic underpinnings, and an 'endophenotype' which has a clear genetic basis. The former might be 'environmental, epigenetic, or multifactorial' in nature (p. 638, Gottesman & Gould, 2003), and the low MZ correlations for WCST performance found in this and other studies (e.g., Campana et al., 1996; Kremen et al., 2007) suggests that it might be a 'biological marker' rather than an 'endophenotype' for executive deficits associated with certain psychiatric disorders.

The present study provided additional data needed to clarify the heritability of WCST and Stroop performance. Strengths of the study include the mixed gender sample of twins, and the representativeness of the sample in terms of race and ethnicity. Limitations of the study include the higher proportion of MZ pairs relative to DZ pairs, and greater number of female pairs relative to male pairs, although these two imbalances are quite typical in volunteer twin samples (Lykken et al., 1987). Nonetheless, gender differences in the estimates of genetic and environmental influence on the WCST and Stroop were not tested and cannot be ruled out. In addition, the ascertainment strategy yielded a sample of mostly young adult twins that were currently in college, and some older adult twins from the community, which is not necessarily representative of all twins in similar age groups. As such, the present findings should be applied with particular caution to populations with lower levels of education. Although the questionnaire and similarity rating method used to determine zygosity has a long history in twin studies, and has demonstrated utility (Lykken et al., 1990), it would have been preferable to use DNA markers to determine zygosity, given the relatively small sample size, and the potential impact of misclassifying pairs. Finally, the sample was larger than most other twin studies of WCST perfor-

mance, but was nonetheless too small to detect modest levels of heritability.

In conclusion, the present study provided data to round out the evidence for moderate heritability on Stroop performance, providing support for at least that criterion of an endophenotype as it applies to the Stroop. The present study was consistent with other twin studies (Campana et al., 1996; Kremen et al., 2007; Nicole & Del Miglio, 1997; see Anokhin et al., 2003 for the exception) in suggesting that WCST does not appear to have a substantial genetic basis, and might not be an appropriate candidate for an endophenotype. Rather, WCST performance could represent a 'biological marker' of a particular executive dysfunction that characterizes certain psychiatric disorders, and, as such, will remain a valuable measure to clinicians and researchers alike.

### Acknowledgment

This article was supported in part by NIMH grant MH067583. I would like to thank Leonardo Bobadilla, Mark Reeves, Lisa James, and talented undergraduate research assistants for their help in collecting the data. I am especially grateful to the twins for their participation.

### References

- Akaike, H. (1987). Factor analysis and the aic. *Psychometrika*, *52*, 317–332.
- Almasy, L., & Blangero, J. (2001). Endophenotypes as quantitative risk factors for psychiatric diseases: Rational and study design. *American Journal of Medical Genetics (Neuropsychiatric Genetics)*, *105*, 42–44.
- Anokhin, A. P., Heath, A. C., & Ralano, A. (2003). Genetic influences on frontal brain function: WCST performance in twins. *Neuroreport*, *14*, 1975–1978.
- Bruder, G. E., Keilp, J. G., Xu, H., Shikhman, M., Schori, E., Gorman, J. M., & Gilliam, T. C. (2005). Catechol-O-Methyltransferase (COMT) genotypes and working memory: Associations with differing cognitive operations. *Biological Psychiatry*, *58*, 901–907.
- Campana, A., Macciardi, F., Gambini, O., & Scarone, S. (1996). The Wisconsin card sorting test (WCST) performance in normal subjects: A twin study. *Neuropsychobiology*, *34*, 14–17.
- de Geus, E. J. C. (2002). Introducing genetic psychophysiology. *Biological Psychology*, *61*, 1–10.
- Doyle, A. E., Faraone, S. V., Seidman, L. J., Willcutt, E. G., Nigg, J. T., Waldman, I. D., Pennington, B. F., Peart, J., & Biederman, J. (2005). Are endophenotypes based on measures of executive functions useful for molecular genetic studies of ADHD? *Journal of Child Psychology and Psychiatry*, *46*, 774–803.
- Egan, M. F., Goldberg, T. E., Kolachana, B. S., Callicott, J. H., Mazzanti, C. M., Straub, R. E., Goldman, D., & Weinberger, D. R. (2001). Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schiz-

- izophrenia. *Proceedings of the National Academy of Sciences*, 98, 6917–6922.
- Fishbein, D. (2000). Neuropsychological function, drug abuse, and violence: A conceptual framework. *Criminal Justice and Behavior*, 27, 139–159.
- Flint, J., & Munafo, M. R. (2007). The endophenotype concept in psychiatric genetics. *Psychological Medicine*, 37, 163–180.
- Giancola, P. R., & Mezzich, A. C. (2000). Neuropsychological deficits in female adolescents with a substance use disorder: Better accounted for by conduct disorder? *Journal of Studies on Alcohol*, 61, 809–817.
- Golden, C. J. (1978). *Stroop Color and Word Test: A manual for clinical and experimental uses*. Wood Dale, IL: Stoelting Company.
- Gottesman, I. I., & Gould, T. D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry*, 160, 636–645.
- Heaton, R. K. (2000). WCST-64: Computer Version for Windows — Research Edition [Computer software]. Odessa, FL: Psychological Assessment Resources, Inc.
- Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual: Revised and Expanded*. Odessa, FL: Psychological Assessment Resources.
- Inoue, K., & Lupski, J. R. (2003). Genetics and genomics of behavioral and psychiatric disorders. *Current Opinion in Genetic Development*, 13, 303–309.
- Johnson, W., Bouchard, T. J. Jr, Segal, N. L., Keyes, M., & Samuels, J. (2003). The Stroop Color-Word Test: Genetic and environmental influences: Reading, mental ability, and personality correlates. *Journal of Educational Psychology*, 95, 58–65.
- Keri, S., & Janka, Z. (2004). Critical evaluation of cognitive dysfunctions as endophenotypes of schizophrenia. *Acta Psychiatrica Scandinavica*, 110, 83–91.
- Kremen, W. S., Eisen, S. A., Tsuang, M. T., & Lyons, M. J. (2007). Is the Wisconsin Card Sorting Test a reliable neurocognitive endophenotype? *American Journal of Medical Genetics (Neuropsychiatric Genetics)*, 144, 403–406.
- Lykken, D. T., Bouchard, T. J. Jr, McGue, M., & Tellegen, A. (1990). The Minnesota Twin Registry: Some initial findings. *Acta Geneticae Medicae et Gemellologiae: Twin Research*, 39, 35–70.
- Lykken, D. T., McGue, M., & Tellegen, A. (1987). Recruitment bias in twin research: The rule of two-thirds reconsidered. *Behavioral Genetics*, 17, 343–362.
- McGue, M., & Bouchard, T. J. Jr. (1984). Adjustment of twin data for the effects of age and sex. *Behavioral Genetics*, 14, 325–343.
- Mills, S., Langley, K., Van den Bree, M., Street, E., Turic, D., Owen, M. J., O'Donovan, M. C., & Thapar, A. (2004). No evidence of association between Catechol-O-Methyltransferase (COMT) Val<sup>158</sup>Met genotype and performance on neuropsychological tasks in children with ADHD: A case-control study. *BMC Psychiatry*, 4, 15.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100.
- Neale, M. C., Xie, G., & Boker, S. M. (1998). *Introduction to the Mx Graphical User Interface*. Richmond, VA: Virginia Commonwealth University, Department of Psychiatry.
- Nicole, S., & Del Miglio, C. (1997). Abstraction skillfulness in monozygotic and dizygotic twin pairs. *Acta Geneticae Medicae et Gemellologiae (Roma)*, 46, 57–67.
- Reuter, M., Peters, K., Schroeter, K., Koebke, W., Lenardon, D., Bloch, B., & Hennig, J. (2005). The influence of the dopaminergic system on cognitive functioning: A molecular genetic approach. *Behavioral Brain Research*, 164, 93–99.
- Rybakowski, J. K., Borkowska, A., Czerski, P. M., Dmitrzak-Weglarz, M., Skibinska, M., Kapelski, P., & Hauser, J. (2006). Performance on the Wisconsin Card Sorting Test in schizophrenia and genes of dopaminergic inactivation (COMT, DAT, NET). *Psychiatry Research*, 143, 13–19.
- Rybakowski, J. K., Borkowska, A., Skibinska, M., Szczepankiewicz, A., Kapelski, P., Leszczynska-Rodziewicz, A., Czerski, P. M., & Hauser, J. (2006). Prefrontal cognition in schizophrenia and bipolar illness in relation to Val66Met polymorphism of the brain-derived neurotrophic factor gene. *Psychiatry and Clinical Neurosciences*, 60, 70–76.
- Stins, J. F., van Baal, G. C. M., Polderman, T. J. C., Verhulst, F. C., & Boomsma, D. I. (2004). Heritability of Stroop and flanker performance in 12-year old children. *BMC Neuroscience*, 5, 49.
- Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: A conceptual view. *Psychological Research*, 63, 289–298.
- Szoke, A., Schurhoff, F., Mathiew, F., Meary, A., Ionescu, S., & Leboyer, M. (2005). Tests of executive functions in first-degree relatives of schizophrenic patients: A meta-analysis. *Psychological Medicine*, 35, 771–781.
- Taylor, J. E., James, L. M., Reeves, M. D., & Bobadilla, L. (2006). The Florida State Twin Registry: Research aims and design. *Twin Research and Human Genetics*, 9, 958–962.
- Taerk, E., Grizenko, N., Amor, L. B., Lageix, P., Mbekou, V., Deguzman, R., Torkaman-Zehi, A., Stepanian, M. T., Baron, C., & Joobar, R. (2004). Catechol-O-Methyltransferase (COMT) Val<sup>108/158</sup>Met polymorphism does not modulate executive function in children with ADHD. *BMC Medical Genetics*, 5, 30.