Published in final edited form as:

Behav Genet. 2012 March ; 42(2): 209–220. doi:10.1007/s10519-011-9502-1.

Phenotypic and Genetic Analyses of the Wisconsin Card Sort

Detre A. Godinez^{1,2}, Naomi P. Friedman³, Soo Hyun Rhee^{2,3}, Akira Miyake², and John K. Hewitt^{2,3}

¹Department of Psychiatry, University of Colorado, Denver, CO

²Department of Psychology and Neuroscience, University of Colorado, Boulder, CO

³Institute for Behavioral Genetics, University of Colorado, Boulder, CO

Abstract

The present study assessed the factor structure and etiology of traditional perseverative and nonperseverative errors, and six narrowly defined errors that occur during the Wisconsin Card Sorting Task (WCST). A computer-administered version of the WCST, designed to maximize the variance in a nonclinical sample, was used. Phenotypic factor analysis and twin models were used to examine the structure and genetic and environmental etiology in 191 monozygotic and 165 dizygotic adolescent twin pairs. Factor analysis did not support the traditional division of errors into perseverative and nonperseverative errors. Heritability of individual indices was small to moderate ($a^2 = .10 - .42$), with varying significance. Estimates of shared environment ($c^2 = .00 - .14$) were not significant. The best fitting multivariate genetic model had one genetic factor, with specific variance and covariance due to nonshared environmental influences. These results suggest that there are common underlying genetic influences on WCST indices, along with index-specific environmental variance that does not correspond to the traditional division between perseverative and nonperseverative errors.

Keywords

executive function; perseveration; twin study; genetics; WCST

Phenotypic and Genetic Analyses of the Wisconsin Card Sort

The Wisconsin Card Sorting Task (WCST), originally developed by Berg (1948) remains one of the most widely used neuropsychological measures of frontal lobe and associated executive functioning. Although the WCST is a commonly used task in both clinical and research settings, the underlying influences in performance on this task are not well understood. By examining the factor structure and etiology of the traditional indices and more narrowly defined errors in a version of the WCST designed to maximize this variance, the present study aims to provide a better understanding of the underlying influences in this task in a nonclinical sample of older adolescents.

The current study had two major goals. The first goal was to assess the covariance and factor structure of WCST errors at the phenotypic level. We examined (a) whether the relationships among the errors were consistent with the traditional classifications of perseverative and nonperseverative errors and (b) whether the WCST is a measure of a more

To whom correspondence should be addressed. Detre Godinez, University of Colorado at Boulder, Institute of Cognitive Science, 594 UCB, Boulder, CO 80309. Phone: 720-771-3824. Fax: (303) 492-7717. godinez@colorado.edu.

None of the authors have any conflict of interest to declare.

general or multiple factors. The second goal of this study was to assess the genetic and environmental structure of the WCST indices. Using monozygotic and dizygotic twin pairs, we examined the degree to which there are shared and specific genetic and environmental influences on the different error types.

In the WCST, participants sort a series of cards according to the shape, color, or number of symbols on the cards. The participants learn the correct sorting principle from only positive or negative feedback. Additionally, the sorting principle switches after a specified number of correct sorts, and they must learn to switch to another category. For example, if the first sorting principle is shape, the participants should sort by shape until the rule changes and negative feedback is received. After the rule changes, they must try the other two sorting principles until they get positive feedback, then sort by this new sorting principle.

The WCST was first designed as a measure of shifting cognitive set and categorization (Grant & Berg, 1948). Later, researchers found that the WCST could discriminate patients with frontal lobe lesions from controls (Milner, 1963). Intuitively, the WCST's complex nature likely taps into multiple lower-level and higher-level processes. This complexity fuels the debate on the use of the WCST as a pure measure of frontal lobe functioning, as there are many brain regions involved in performance (Andres, 2003; Mountain & Snow, 1993; Nyhus & Barcelo, 2009; Reitan & Wolfson, 1994). On the other hand, the WCST's complex nature and its association with prefrontal activity also make the WCST a potentially good measure of general executive function (Buchsbaum et al., 2005; Demakis, 2003).

The continued use of the WCST in clinical assessments suggests its value in patient populations (Rabin et al., 2005). Performance on the WCST has been associated with several disorders. Although the effect sizes are small to moderate, several meta-analyses have reported that negative symptoms and disorganization symptoms of schizophrenia were associated with more perseverative errors in the WCST (Li, 2004; Nieuwenstein *et al.*, 2001; Szoke *et al.*, 2005). Furthermore, it was found that the errors on the WCST could differentiate the relatives of schizophrenic patients and controls (Szoke *et al.*, 2005). Meta-analyses have also concluded that the effect sizes of the relationship between the WCST and ADHD are small but significant (Pennington & Ozonoff, 1996; Willcutt *et al.*, 2005). In another meta-analytic review, Morgan and Lilienfeld (2000) found evidence for a negative correlation between the WCST performance and antisocial behavior. The WCST was also found to have a pooled medium effect size in a systematic review and meta-analysis of eating disorders and set shifting tasks (Roberts et al., 2007). Other disorders that may be associated with WCST performance include unipolar depression and substance use disorders (Giancola & Mezzich, 2003; Moritz et al., 2002). These meta-analyses, especially the meta-analysis on the relatives of individuals with schizophrenia, suggest the WCST is a useful measure of a cognitive deficit or deficits common to several disorders.

Despite continued use of the WCST, many studies have lacked the power to understand the underlying influences of the WCST. Genetic studies in particular have resulted in inconsistent findings regarding the heritability of WCST performance (Anokhin et al., 2010; Anokhin et al., 2003; Campana et al., 1996; Chou et al., 2010; Kremen et al., 2007; Taylor, 2007). Table 1 shows a summary of results from heritability studies of the WCST, including sample size and errors assessed. Small samples and the use of different indices and versions of the WCST have likely contributed to these inconclusive results. Four of the six studies have at least one group with a sample size under fifty twin pairs, which is a relatively small sample for heritability studies. The errors assessed also vary widely from one study to the next. Although perseverative errors is an error type assessed in all of the studies, most studies do not report how these errors were defined and, in the few studies that do report error definitions, perseverative error definitions vary across studies. The inconsistencies

across these studies show that although the WCST is a commonly used task, a better understanding of the genetic and environmental influences on these performance indices is needed.

There are several versions of the WCST and several ways to score it. Perseverative errors (incorrect sorting according to the previously correct category) is the most popular index of the WCST, with number of categories achieved as the second most widely used index. Traditional perseverative errors can be defined either as errors that occur after a category switch in which the participant continues to sort by the previously correct category, or more specifically as errors that occur after a person has successfully switched to a new category and then resorts back to sorting by the previously correct category. Nonperseverative or random errors are usually defined as any other type of error. For example, given there are three sorting categories (shape, color, and size), a nonperseverative error would occur when the participant switches to an incorrect category (other than the previously correct category) either because the participant fails to understand the rule or he/she loses track of the current category. Other possibilities for error types that often overlap with perseverative or nonperseverative errors include perseverative responses, failure to maintain set errors, conceptual level response, trials to complete first category, learning to learn, total errors, and percentage of errors (Heaton et al., 1993; Kimberg et al., 1997)

To complicate matters, after a category switches and negative feedback is given, necessary and functional errors are possible in the first and second trials for participants to determine the new sorting principle. These errors are referred to as efficient errors (Barcelo, 1999; Barcelo & Knight, 2002). Although the correct category can be found by chance in the first or second trial after a category switch, it is not until the third trial that the correct categorization could reliably be found since there are three categories to choose from. In an electrophysiological study (Barcelo, 1999) and a neuropsychological study of comparing patients with lateral prefrontal lesions and a control sample (Barcelo & Knight, 2002), Barcelo and colleagues found that efficient errors should not be defined as perseverative or nonperseverative errors and that including them in the nonperseverative category obscured the distinction between patients and controls.

The Present Study

The present study takes into account the wide range of error definitions and classifies the WCST errors into the six distinct, narrowly defined error types listed in Table 2. A systematic review of the literature of the WCST suggested that errors definitions varied from one study to the next and that many of the errors were often a combination of a couple of error types. Therefore, in the present study, errors were defined distinctly based on the most narrowly defined and distinct error definitions. This review of the literature also demonstrated a conceptual distinction, such that errors could be defined more narrowly based on whether the errors occurred before or after the participant correctly sorted. By assessing the variance, covariance, and factor structure among these distinct errors, one can understand their relations better.

In the present study, traditional perseverative errors were divided into two categories: perseverative errors before a correct sort and perseverative errors after a correct sort. Traditional nonperseverative errors were also divided into categories: efficient errors, search errors (nonperseverative errors before a correct sort), breaking set errors (nonperseverative errors after a correct sort), and repeated nonperseverative errors (two or more nonperseverative errors in a row). Search errors, which occur before a correct sort, are thought to reflect the participant having difficulty finding the correct category. Breaking set errors occur when a person has correctly sorted and therefore has received the positive

feedback about the correct sorting principle but breaks set either because he/she is distracted or unable to use the available information to guide sorting.

We defined the errors more narrowly to gain insight into some of the questions regarding the WCST. One such question is whether the established division of the WCST errors into perseverative and nonperseverative errors reflects underlying covariance. If so, results of factor analyses should suggest that perseverative errors before and after a correct sort should load on one factor and nonperseverative errors should load on another factor. A second related question is whether the errors are measures of multiple abilities or a measure of a single or more general executive function. Assessing the covariance among the new narrowly defined errors can provide insight regarding this question.

The present study also assessed the etiology of the WCST indices by examining the magnitude of genetic and environmental influences on the WCST indices. In addition to obtaining the magnitude of genetic and environmental influences for each error type, we used a multivariate approach to assess the genetic and environmental influences on covariance between different errors. We tested whether this covariance is explained by a general common factor, which supports the use of the WCST as a measure of a single general executive function factor, or multiple factors, which suggests that the WCST errors index multiple abilities.

Method

Participants

Participants were twin pairs from the Colorado Longitudinal Twin Study, a twin sample obtained from the general population (see Rhea et al., 2006). A total of 797 individual twins were assessed at approximately age 17 ($M = 17.3$, $SD = 0.6$, Range = 15.8 to 20.1). From this sample, 45 were excluded from analyses for the following reasons. Six individuals were colorblind. Eighteen participants did not complete the WCST portion of the assessment. For twenty-one individuals, it was unclear as to whether they understood the instructions because they did not achieve at least 5 categories. The remaining 752 individuals included 40 singletons, 191 monozygotic (MZ) pairs, and 165 same-sex dizygotic (DZ) twin pairs. Females represented 104 MZ pairs and 85 DZ pairs and males represented 87 MZ pairs and 80 DZ pairs.

Zygosity Determination

Interviewers rated the twins at multiple time points on a nine-item questionnaire of physical characteristics (Nichols & Bilbro, 1966). Additionally, zygosity was determined by concordance of at least 11 highly informative short tandem repeat polymorphisms genotyped with standard polymerase chain reaction methods and ABI 377 technology. Reevaluations and further genotyping were conducted when there were discrepancies between zygosity derived from the two methods.

Procedures and Measures

Written informed assent (from minor participants) and consent (from guardians of minor participants) were obtained from all participants. Participants were paid \$50 for their involvement in the study, which included other cognitive tasks not examined in the current analyses.

A computerized speeded version of the WCST, implemented by Kimberg et al. (1997), was used. This task maximizes variance and can be used in a nonclinical sample. For each trial, a target card containing a set of stimuli that could be categorized by color (red, green, blue, or

yellow), shape (cross, circle, square, or star), or number (1, 2, 3, or 4) was presented. Participants sorted each card into a pile based on its match to four cards that appeared at the bottom of the screen. These four cards mapped with four keys on the keyboard (a, s, k, & l) that were separated from other keys. Participants received positive feedback (“right”), negative feedback (“wrong”), or timeout feedback (when 3 seconds from the time the card appeared had lapsed without a response). Eight consecutive correct categorizations (no timeouts) were needed before the category switched. The maximum number of trials was 288.

Measures included the two traditional scores, perseverative errors and nonperseverative errors. In our sample and in most nonpatient samples, the traditional number of categories achieved is skewed because most participants complete all 15 possible categories. We therefore used the number of trials to achieve 15 categories as a measure of overall achievement (or the maximum number of trials, 288, if they did not achieve all 15 categories). In addition to these traditionally defined errors, we narrowed the definitions of perseverative and nonperseverative errors and created six distinct error types shown in Table 2. Efficient errors are only defined as such if a correct sort is made in the next trial or if the error in the next trial leads to a correct sort in the third trial. Although an error in the second trial that leads to a correct sort in the third trial is also efficient, we only counted the first trial because chance dictates whether a person makes one or two efficient errors per trial.

Data Analyses

Data transformation—All measures except efficient errors and number of trials showed skewed or kurtotic distributions and were log transformed to better approximate normality. Transformation of the data resulted in acceptable kurtosis and skewness (see Table 3). Split-half reliability estimates were obtained in SPSS using odd versus even categories.

Phenotypic factor analyses—Phenotypic factor analyses were performed in Mplus. Exploratory factor analysis (EFA) was performed using oblique and orthogonal rotation. Eigenvalues over 1 were used as criteria for factor numbers selection. The overall fit of the model was estimated using χ^2 , Bentler’s comparative fit index (CFI), Akaike’s information criterion ($AIC = \chi^2 - 2df$) and root-mean-square error of approximation (RMSEA). A nonsignificant χ^2 , lower AIC, RMSEA < .06, and CFI > .95 suggest that the model fits the data well (Hu & Bentler, 1998). Nonindependence of twin pairs was addressed by using an Mplus option (Type = complex, Cluster = family) that estimates parameters, standard errors, and the Yuan-Bentler χ^2 that are robust to nonindependence.

Univariate genetic models—Univariate analysis of twin data was conducted on the traditional and narrowly defined indices of the WCST using Mx (Neale et al., 1999). A standard univariate twin model was used in which the variance of the phenotypes are decomposed into additive genetic, dominant genetic, shared environmental, and nonshared environmental variance, which includes error variance (Neale & Maes, 1999). In this model, MZ twins share 100% of additive and dominant genetic effects, DZ twins share 50% of additive genetic effects and 25% of dominant genetic effects, and both MZ and DZ twins share 100% of shared environment effects and 0% of nonshared environmental effects.

A limitation of the twin-only design is that dominant genetic and shared environmental influences cannot be estimated simultaneously because they are confounded. Therefore, we tested the ACE (including additive genetic, shared environmental, and nonshared environmental influences), ADE (including additive genetic, dominant genetic, and nonshared environmental influences), and AE (including additive genetic and nonshared environmental influences) models. Models that decompose the variance into only dominant

genetic effects and nonshared environmental effects (i.e., DE) are biologically implausible and hence were not tested.

Multivariate genetic models—Multivariate analyses were also conducted in Mx. The same principles apply, although this analysis decomposes the covariance among measures (in addition to the variance) into additive genetic, dominant genetic, shared environmental, and nonshared environmental influences. Overall fit of the model was assessed by comparing minus twice the log likelihood ($-2LL$) of the saturated model and the $-2LL$ of the model in question. The difference between $-2LL$ approximates a χ^2 distribution. Root-mean-square error of approximation (RMSEA) was obtained by comparing the fit of the model to the saturated and null model. A nonsignificant χ^2 , $RMSEA < .06$ and a lower AIC ($AIC = \chi^2 - 2df$) suggests a better fitting model. The most parsimonious model is found by dropping parameters that do not contribute to the fit of the model. The χ^2 difference test ($\Delta\chi^2$) was used to test the fit of alternative models.

Results

Descriptive Analyses

Table 3 displays the descriptive statistics including the split-half reliability estimates (based on odd-even categories) for the traditional and more narrowly defined errors. Split-half coefficient reliabilities are moderate to high for the more narrowly defined errors (range .63–.86), with the exception of efficient errors (.37). The small range for efficient errors (a minimum of 0 errors and a maximum of 13 errors) may have led to a low reliability. Consistent with Bowden et al.'s findings (1998), reliabilities for the traditionally defined errors (.21–.63) were lower than those of the more narrowly defined errors, suggesting that the latter are more consistent across trials.

All WCST errors, both traditional and narrowly defined, were tested for sex and age main effects. Due to the narrow age range, no main effects of age were found and therefore no corrections were made for age. Small but significant sex effects were found for perseverative errors before a correct sort and search errors ($\beta_{perBCS} = -.11$, $p = .005$; $\beta_{search} = -.12$, $p = .002$). Results did not differ between analyses performed on raw data and analyses performed on residual data; the following results are for analyses performed on residual scores after the effects of sex on perseverative before a correct sort and search errors have been regressed out.

If traditional error definitions reflect underlying covariance: 1) errors typically defined as perseverative (perseverative errors before and after a correct sort) should correlate highly with one another; 2) errors typically defined as nonperseverative (search, breaking set, repeated nonperseverative, and efficient errors) should correlate highly with one another; and 3) perseverative errors and nonperseverative errors should be minimally correlated. As shown in table 4, the data do not support these predictions. First, the correlation between the traditional nonperseverative errors and perseverative errors was high ($r = .69$). Second, the traditional perseverative error definition is a composite of perseverative errors before and after a correct sort; however, the correlation between these two errors was relatively low ($r = .30$). Third, among the errors that are traditionally defined as nonperseverative, correlations varied from .21 to .50, with negative correlations between efficient errors and the other errors.

Phenotypic Factor Analyses

Table 5 displays results from confirmatory and exploratory factor analyses. To further examine whether the traditional definitions (i.e., perseverative and nonperseverative)

adequately describe the data, we tested a two-factor confirmatory factor analysis model. In this model, errors that are typically defined as perseverative (perseverative errors before and after a correct sort) loaded on one factor and errors typically defined as nonperseverative (search, breaking set, repeated nonperseverative, and efficient errors) loaded on a second factor. This model fit the data poorly ($\chi^2(8) = 769.58$, AIC = 753.58, CFI = .64, RMSEA = .36).

To examine how these errors should be categorized, we used exploratory and confirmatory factor analysis. Results from exploratory factor analysis suggested a two-factor model, $\chi^2(4) = 67.02$, fit the data better than a one-factor model, $\chi^2(9) = 755.76$ (eigenvalues for the EFA: $\lambda_1 = 3.26$, $\lambda_2 = 1.33$, $\lambda_3 = 0.65$). Interestingly, the two factors were fairly distinct with the errors occurring before a correct sort loading highly on the first factor and errors occurring after a correct sort loading highly on the second factor. Confirmatory factor analysis suggested that a 2 factor model, in which perseverative errors before a correct sort, and search errors load on the first factor and perseverative errors after a correct sort, breaking set errors, and repeated nonperseverative errors load on the second factor but where efficient errors loaded negatively and significantly on both factors, provided the best fit to the data.

In summary, correlations and results from exploratory and confirmatory factor analyses provide evidence against the traditional grouping of perseverative and nonperseverative errors on a phenotypic level. Rather, the exploratory and confirmatory factor analyses suggested two factors, one for errors before a correct sort and the other for errors after a correct sort. Hence we used this suggested factor structure in the multivariate genetic models.

Genetic Analysis

Univariate twin models—We examined the heritability of both the traditionally defined errors and the more narrowly defined errors. In all cases, the fit of the ACE and ADE models were similar; therefore, only ACE models are shown. Table 6 shows MZ/DZ correlations. Parameter estimates and fit statistics from the univariate ACE twin models are shown in Table 7. For most errors, the magnitudes of shared environmental influences were estimated at or close to zero. Repeated nonperseverative errors ($c^2 = .14$) showed the largest estimate; however, confidence intervals for all shared environment estimates included zero. Small to moderate heritability estimates (.10 – .42) were found for the traditionally and narrowly defined errors. Heritability estimates were significant for traditional nonperseverative errors, number of trials, search errors, breaking set errors, and efficient errors.

Multivariate twin models—Figure 1 shows the alternative multivariate models tested to examine whether the different errors share common genetic and environmental influences. Model A is the Cholesky model (Neale & Maes, 1999), which is an unrestricted model that fully captures the covariance among the six errors. Both AE and ACE Cholesky models were assessed, but only the AE model is shown for the sake of simplicity. Model B is a two-factor genetic model based on the results of the exploratory factor analysis (i.e., perseverative errors after a correct sort, breaking set errors, efficient errors, and repeated nonperseverative errors load on one factor and perseverative errors before a correct sort, efficient errors, and search errors load on another factor). This model allows for the two factors to be correlated with specific genetic influences on each individual variable. Model C is a one-factor genetic model with specific genetic influences and Model D is a one-factor genetic model that constrains the genetic covariance between the variables to a single

common factor. All of these models allow for the nonshared environmental effects to follow the form of the Cholesky, which allows for error variance and covariance.

As shown in Table 8, the fit of the Cholesky AE model was not significantly worse than the Cholesky ACE model, indicating nonsignificant C parameters ($\Delta\chi^2_{21} = 5.09$, $p = 1.0$). Although not shown, the shared environment (C parameters) could be dropped in all models. The correlated two-genetic-factor model (Model B), which tested the fit of a two-factor model based on the results of exploratory factor analysis, also did not result in a significant decrement of fit ($\Delta\chi^2_7 = 8.76$, $p = .27$). The genetic correlation between these factors was estimated at .82 and when constrained to 1.0 did not result in a significant decrement of fit ($\Delta\chi^2_1 = 2.8$, $p = .10$). Although not shown, the correlation for the two-factor model based on the standard division of perseverative and nonperseverative errors was estimated at 1.0. The one-genetic-factor without trait-specific genetic influences (Model D) was the most parsimonious model (Model C with trait-specific genetic influences, $\Delta\chi^2_2 = 2.8$, $p = .25$; Model D without trait-specific genetic influences, $\Delta\chi^2_6 = 8.2$, $p = .23$). Figure 2 shows the factor loadings for this model, with the nonshared environmental paths that could be dropped shown in grey, dotted lines ($\Delta\chi^2_6 = 9.13$, $p = .17$). Constraining the nonshared environmental covariance such that only trait specific variance remained resulted in a large decrement of fit compared to Model D ($\Delta\chi^2_{15} = 1120.44$, $p = .00$). The finding that the best-fitting, most parsimonious model was the one-genetic-factor model without trait-specific genetic loadings suggests that there is one set of significant genetic influences shared in common by all of the errors and that specificity in the errors is due to nonshared environmental influences.

Discussion

The goals of the present study were 1) to assess the covariance and factor structure of the WCST indices to obtain a better understanding of the relations among the error types and 2) to assess the etiology (genetic and environmental) of the variance and covariance of the WCST indices. Multivariate individual differences analyses in our study suggest that there are common underlying genetic influences on WCST indices. The definitions of the errors, however, play a role in the reliability, variance, and nonshared environmental or error covariance.

Phenotypic Factor Structure

During the WCST, it is necessary to out-compete or bias the previously reinforced sorting principle and respond to the changing environmental demand. Perseverative errors are thought to occur when the participant gets stuck in a set and fails to shift, whereas the basic mechanisms involved in nonperseverative errors may be a failure to maintain the task set and the inability to ignore distractions and interference (Barcelo, 1999; Barcelo & Knight, 2002). Although these distinctions seem intuitively valid, our results do not support them. Differences in reliabilities, phenotypic correlations, and results of the confirmatory and exploratory factor analyses suggest that the errors on the WCST should be defined more narrowly.

Perseverative errors before a correct sort and perseverative errors after a correct sort, which are typically defined as perseverative, did not correlate with each other highly. The correlations among the errors typically defined as nonperseverative also varied against expectations such that some correlated better with errors typically defined as perseverative than they did with each other. Additionally, the reliabilities of the traditionally defined errors were low to moderate, whereas the reliabilities of the more narrowly defined errors were higher. This result is consistent with low reliabilities for traditional WCST errors reported by others (Bowden et al., 1998). The negative correlations for efficient errors and the other

WCST indices, which are consistent with the findings of Barcelo and colleagues (1999; 2002), also support the importance of defining errors distinctly.

Etiology of the Factor Structure

Although a handful of studies have assessed the heritability of the WCST, the results have varied considerably. The results of our univariate twin analyses suggest that all of the errors, traditionally or narrowly defined, are modestly to moderately heritable ($h^2 = .10-.42$) although with varied significance. The magnitudes of shared environmental influences were low or estimated at zero for most errors. Further research will be important to clarify nonadditive genetic effects, which could also be contributing to inconsistent results.

The etiology of the covariance among the WCST indices was also one of the questions this study examined. The results from the genetic multivariate models suggest that the narrowly defined errors share the same underlying genetic influences and what differentiate them are nonshared environmental factors. Although a correlated two-factor model (Model B), based on the exploratory factor analysis, provided an adequate fit to the data, the most parsimonious model was a one-genetic-factor model (Model D). This result is consistent with those from Greve et al. (2005), who found that although a three-factor model also fit the data, the evidence for the additional factors was less stable and likely due to error variance. Interestingly, when the nonshared environmental or error covariance was constrained such that nonsignificant paths were set to zero, the first two factors reflected results from our exploratory factor analysis. Thus, despite the intuitive distinction between perseverative and nonperseverative errors, differences among indices in this sample may be more accurately described as errors occurring before a correct sort and errors occurring after a correct sort.

Our study is comparable with the largest WCST twin studies to date; however, an increased sample size would have allowed us to make better distinctions among the alternative multivariate models and test for sex differences. MZ/DZ correlations suggest possible nonadditive genetic effects and sex differences. It would be valuable to test for these effects in samples with the necessary power. Another qualification of the current results is that although using a community sample has its advantages, these results cannot be generalized to clinical samples. A study replicating the examination of the narrowly defined errors in a clinical sample would be a valuable contribution to the understanding of the WCST.

Implications and Conclusions

In our attempt to understand underlying influences in cognitive ability, it is important to define the variables of interest in ways that reflect these influences while not masking these effects. Individual differences research provides an opportunity to assess the underlying influences on task performance. Our results suggest that the traditional distinction of the WCST's errors into perseverative and nonperseverative errors did not capture patterns of individual differences in a nonclinical sample. Rather, our results suggest that the narrowly defined and distinct errors share the same underlying genetic influences. To the extent that this common genetic factor reflects a biologically-based executive function, this result is consistent with the conclusion that the different error types reflect a common general EF factor instead of multiple executive functions (Miyake et al., 2000; Friedman et al., 2008).

Although the variance in the narrowly defined errors could be attributed to the same underlying genetic influences, specific genotypes may contribute to the variance of these errors differently. The definitions of the WCST errors are likely to impact studies assessing molecular genetic associations with performance. The narrowly defined errors were more reliable, and nonperseverative errors before and after a correct sort were consistently

heritable, which may suggest a need to look beyond traditional perseverative errors when assessing individual differences in nonclinical samples. Further exploration of these errors may increase power to detect differences when examining associations with specific genes and may lead to a better understanding of the genetic and environmental influences on the WCST.

Acknowledgments

This research was supported by National Institute of Health grants MH063207, DA0111015, MH079485, T32HD002789, and T32MH15442.

We thank Marie Banich for support and assistance with the manuscript and Sally Ann Rhea for coordination of data collection.

References

- Andres P. Frontal cortex as the central executive of working memory: Time to revise our view. *Cortex*. 2003; 39:871–895. [PubMed: 14584557]
- Anokhin AP, Golosheykin S, Grant JD, Heath AC. Developmental and genetic influences on prefrontal function in adolescents: A longitudinal twin study of WCST performance. *Neurosci Lett*. 2010; 472:119–122. [PubMed: 20132870]
- Anokhin AP, Heath AC, Ralano A. Genetic influences on frontal brain function: WCST performance in twins. *Neuroreport*. 2003; 14:1975–1978. [PubMed: 14561932]
- Barcelo F. Electrophysiological evidence of two different types of error in the Wisconsin Card Sorting Test. *Neuroreport*. 1999; 10:1299–1303. [PubMed: 10363943]
- Barcelo F, Knight RT. Both random and perseverative errors underlie WCST deficits in prefrontal patients. *Neuropsychologia*. 2002; 40:349–356. [PubMed: 11684168]
- Berg EA. A simple objective technique for measuring flexibility in thinking. *J Gen Psychol*. 1948; 39:15–22. [PubMed: 18889466]
- Buchsbaum BR, Greer S, Chang WL, Berman KF. Meta-analysis of neuroimaging studies of the Wisconsin Card-Sorting task and component processes. *Hum Brain Mapp*. 2005; 25:35–45. [PubMed: 15846821]
- Bowden SC, Fowler KS, Bell RC, Whelan G, Clifford CC, Ritter AJ, Long CM. The reliability and internal validity of the Wisconsin Card Sorting Test. *Neuropsychol Rehabil*. 1998; 8:243–254.
- Campana A, Macciardi F, Gambini O, Scarone S. The Wisconsin Card Sorting Test (WCST) performance in normal subjects: A twin study. *Neuropsychobiology*. 1996; 34:14–17. [PubMed: 8884753]
- Chou LN, Kuo PH, Lin CC, Chen WJ. Genetic and environmental influences on the Wisconsin Card Sorting Test performance in healthy adolescents: A twin/sibling study. *Behav Genet*. 2010; 40:22–30. [PubMed: 19816765]
- Demakis GJ. A meta-analytic review of the sensitivity of the Wisconsin Card Sorting Test to frontal and lateralized frontal brain damage. *Neuropsychology*. 2003; 17:255–264. [PubMed: 12803431]
- Friedman NP, Miyake A, Young SE, DeFries JC, Corley RP, Hewitt JK. Individual differences in executive functions are almost entirely genetic in origin. *J Exp Psycho Gen*. 2008; 137:201–225.
- Giancola PR, Mezzich AC. Executive functioning, temperament, and drug use involvement in adolescent females with a substance use disorder. *J Child Psychol Psych*. 2003; 44(6):857–866.
- Grant DA, Berg EA. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a weigl-type card-sorting problem. *J Exp Psychol*. 1948; 38:404–411. [PubMed: 18874598]
- Greve KW, Stickler TR, Love JM, Bianchini KJ, Stanford MS. Latent structure of the Wisconsin Card Sorting Test: A confirmatory factor analytic study. *Arch Clin Neuropsychol*. 2005; 20:355–364. [PubMed: 15797171]
- Heaton, RK.; Chelune, GJ.; Talley, JL.; Kay, GG.; Curtis, G. Wisconsin card sorting test manual: Revised and expanded. Odessa, FL: Psychological Assessment Resources; 1993.

- Hu L, Bentler PM. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychol Methods*. 1998; 3:424–453.
- Kimberg DY, D'esposito M, Farah MJ. Effects of bromocriptine on human subjects depend on working memory capacity. *Neuroreport*. 1997; 6:3581–3585. [PubMed: 9427330]
- Kremen WS, Eisen SA, Tsuang MT, Lyons MJ. Is the Wisconsin card sorting test a useful neurocognitive endophenotype? *Am J Med Genet. Part B, Neuropsychiatric Genetics*. 2007; 144:403–406.
- Li CS. Do schizophrenia patients make more perseverative than non-perseverative errors on the Wisconsin Card Sorting Test? A meta-analytic study. *Psychiat Res*. 2004; 129(2):179–190.
- Milner B. Effects of different brain lesions on card sorting. *Arc Neurol*. 1963; 9:101–110.
- Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychol*. 2000; 41:49–100.
- Morgan AB, Lilienfeld SO. A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. *Clin Psychol Rev*. 2000; 20(1):113–136. [PubMed: 10660831]
- Moritz S, Birkner C, Kloss M, Jahn H, Hand I, Haasen C, et al. Executive functioning in obsessive-compulsive disorder, unipolar depression, and schizophrenia. *Arch Clin Neuropsych*. 2002; 17(5): 477–483.
- Mountain MA, Snow WG. Wisconsin Card Sorting Test as a measure of frontal pathology: A review. *Clin Neuropsychol*. 1993; 7:108–118.
- Neale, M.; Boker, S.; Xie, G.; Maes, H. Mx: Statistical Modeling. Richmond, VA: Department of Psychiatry, Medical College of Virginia; 1999.
- Neale, MC.; Maes, HH. Methodology for genetic studies of twin and families. Dordrecht, Netherlands: Kluwer Academic Publishers B.V.; 1999.
- Nichols RC, Bilbro WC Jr. The diagnosis of twin zygosity. *Acta Genet Stat Med*. 1966; 16:265–275. [PubMed: 4959152]
- Nieuwenstein MR, Aleman A, de Haan EH. Relationship between symptom dimensions and neurocognitive functioning in schizophrenia: a meta-analysis of WCST and CPT studies. Wisconsin Card Sorting Test. *J Psychiat Res*. 2001; 35(2):119–125. [PubMed: 11377441]
- Nyhus E, Barcelo F. The Wisconsin Card Sorting Test and the cognitive assessment of prefrontal executive functions: A critical update. *Brain Cogn*. 2009; 71:437–451. [PubMed: 19375839]
- Pennington BF, Ozonoff S. Executive functions and developmental psychopathology. *J Child Psychol Psc*. 1996; 37(1):51–87.
- Rabin LA, Barr WB, Burton LA. Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Arch Clin Neuropsychol*. 2005; 20:33–65. [PubMed: 15620813]
- Reitan RM, Wolfson D. A selective and critical review of neuropsychological deficits and the frontal lobes. *Neuropsychol Rev*. 1994; 4:161–198. [PubMed: 7881456]
- Rhea SA, Gross AA, Haberstick BC, Corley RP. Colorado twin registry. *Twin Res Hum Genet*. 2006; 9(6):941–949. [PubMed: 17254434]
- Roberts ME, Tchanturia K, Stahl D, Southgate L, Treasure J. A systematic review and meta-analysis of set-shifting ability in eating disorders. *Psychol Med*. 2007; 37(8):1075–1084. [PubMed: 17261218]
- Szoke A, Schurhoff F, Mathieu F, Meary A, Ionescu S, Leboyer M. Tests of executive functions in first-degree relatives of schizophrenic patients: a meta-analysis. *Psychol Med*. 2005; 35(6):771–782. [PubMed: 15997598]
- Taylor J. Heritability of Wisconsin Card Sorting Test (WCST) and Stroop Color-Word Test performance in normal individuals: Implications for the search for endophenotypes. *Twin Res Hum Genet*. 2007; 10:829–834. [PubMed: 18179394]
- Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. *Biol Psychiat*. 2005; 57(11):1336–1346. [PubMed: 15950006]

Fig. 1.

Multivariate models used to test the relationship among the narrowly defined WCST errors. A, additive genetic influences; C, shared environmental influences; E, nonshared environmental influences; Per BCS, perseverative errors before a correct sort; Per ACS, perseverative errors after a correct sort; Repeated, repeated nonperseverative errors; Br Set, breaking set errors; Efficient, efficient errors; Search, search errors. Cholesky ACE not shown for simplicity.

Fig. 2.

The best fitting model: one genetic factor, no A specifics, cholesky E. Path coefficients in grey could be dropped without resulting in a significant decrement in fit. A, additive genetic influences; E, nonshared environmental influences; BCS, before a correct sort; ACS, after a correct sort.

Table 1**Twin Studies of the Wisconsin Card Sorting Task**

Author, Year	Sample	Errors Examined	Heritability	Shared Environment
Campana et al., 1996	15 MZ, 9 DZ, and 15 unrelated pairs	Total errors, perseverative errors, number of stages completed	NS; NS after controlling for education	NS
Anokhin et al., 2003	58 MZ and 25 DZ female twin pairs	Failure to maintain set, number of categories completed, number & percentage of total errors, perseverative errors, perseverative responses	Failure to maintain set & categories completed NS; other errors 31–46%.	NS
Kremen et al., 2007	170 MZ and 160 DZ male twin pairs	Number of trials, total errors, perseverative errors, number of categories completed, number of trial to complete first category, failure to maintain set	NS	NS
Taylor 2007	80 MZ (22 male) and 29 DZ (4 male) same sex twin pairs	Number of categories completed, total errors, perseverative errors, perseverative responses, nonperseverative errors	NS	NS
Anokhin et al., 2010	166 MZ and 201 DZ pairs	Perseverative errors	19% age 12 49% age 14 females only	NS age 12 34% age 14 males only
Chou et al., 2010	257 MZ, 93 same-sex and opposite sex DZ and 47 same-sex sib-pairs	Total errors, nonperseverative errors, perseverative errors, perseverative responses, categories achieved, conceptual level response, trials to complete first category, learning to learn, failure to maintain set	NS	0–38%

MZ, monozygotic; DZ, dizygotic; NS, nonsignificant

Table 2

Description of Narrowly Defined Errors

Error Type	Traditionally Defined as	Occurrence	Description
Perseverative errors BCS	Perseverative errors	Before a correct sort	A participant chooses the previously correct category despite receiving current negative feedback
Perseverative errors ACS	Perseverative errors	After a correct sort	A participant successfully switches to a new category and correctly sorts, then reverts back to the most recently completed category
Efficient errors	--	Within the first couple trials after a category switch	Errors necessary to eliminate the incorrect categories made after a category switch that lead to a correct sort
Search errors	Nonperseverative errors	Before a correct sort	After a category switch, the participant chooses the incorrect but <i>not</i> previously correct category
Breaking set errors	Nonperseverative errors	After a correct sort	After a correct sort, a participant breaks set and chooses another category, and that category was not the previously reinforced category
Repeated nonperseverative errors	Nonperseverative errors	2 or more nonperseverative Errors in a row	The participant repeatedly chooses the wrong category (2 or more errors) despite negative feedback, but the wrong category is not the previously reinforced category

BCS, before a correct sort; ACS, after a correct sort.

Table 3

Descriptive Table of the WCST Indices

Measure	Mean	SD	Min	Max	Skew ^a	Kurtosis ^a	Reliability ^a
<u>Traditional indices</u>							
Perseverative errors	27.74	16.33	2	122	-0.26	-0.16	0.63
Nonperseverative errors	22.18	10.34	1	66	-0.11	0.87	0.32
Number of trials	228.46	39.96	148	288	0.11	-1.24	0.21
<u>Narrowly defined errors</u>							
Perseverative errors BCS	15.86	10.78	1	105	-0.33	0.27	0.86
Perseverative errors ACS	10.56	9.41	0	66	-0.17	0.04	0.83
Repeated nonperseverative errors	4.28	4.02	0	24	-0.19	-0.54	0.79
Breaking set errors	8.66	6.94	0	39	-0.30	-0.19	0.74
Efficient errors	5.53	2.74	0	13	0.10	-0.54	0.37
Search errors	20.67	7.00	3	50	-0.31	0.82	0.63

^aThese analyses were performed on the transformed data

BCS, before a correct sort; ACS, after a correct sort. n = 752

Table 4

Correlations of WCST Indices

Measure	1	2	3	4	5	6	7	8	9
<u>Traditional indices</u>									
1. Perseverative errors	-								
2. Nonperseverative errors	0.69	-							
3. Number of trials	0.85	0.78	-						
<u>Narrowly defined errors</u>									
4. Perseverative errors BCS	0.81	0.40	0.56	-					
5. Perseverative errors ACS	0.74	0.65	0.80	0.30	-				
6. Repeated nonperseverative errors	0.49	0.78	0.52	0.26	0.45	-			
7. Breaking set errors	0.66	0.80	0.84	0.31	0.77	0.50	-		
8. Efficient errors	-0.77	-0.53	-0.64	-0.75	-0.46	-0.41	-0.45	-	
9. Search errors	0.53	0.50	0.44	0.65	0.17	0.37	0.21	-0.65	-

Note. All correlations are significant. BCS, before a correct sort; ACS, after a correct sort. n=752

Table 5

Phenotypic Factor Analyses-Narrowly Defined Errors

Model	χ^2	AIC	RMSEA	CFI	df
CFA 2 Factor-Per and Nonper	769.58	753.58	0.36	0.64	8
EFA 1 Factor	755.76	737.76	0.33	0.68	9
EFA 2 Factor	67.02	59.02	0.15	0.97	4
CFA 2 Factor-BCS and ACS	104.25	90.25	0.14	0.95	7

χ^2 , chi-square; AIC, Akaike's Information Criterion; RMSEA, root-mean-square error of approximation; CFI, comparative fit index; df, degrees of freedom; CFA, confirmatory factor analysis; Per and Nonper, traditional perseverative and nonperseverative factor model; EFA, exploratory factor analysis; BCS and ACS, before a correct sort and after a correct sort factor model; Eigenvalues for the EFA $\lambda_1 = 3.26$, $\lambda_2 = 1.33$, $\lambda_3 = 0.65$, $n = 752$.

Table 6**MZ DZ Twin Correlations**

Measure	rMZ	CI	rDZ	CI
<u>Traditional indices</u>				
Perseverative errors	0.45	(.35, .54)	0.27	(.11, .41)
Nonperseverative errors	0.40	(.27, .50)	0.09	(.00, .23)
Number of trials	0.46	(.35, .55)	0.27	(.12, .40)
<u>Narrowly defined errors</u>				
Perseverative errors BCS	0.30	(.18, .41)	0.20	(.05, .33)
Perseverative errors ACS	0.33	(.20, .44)	0.19	(.03, .33)
Repeated nonperseverative errors	0.23	(.10, .35)	0.20	(.04, .34)
Breaking set errors	0.38	(.26, .49)	0.13	(.00, .27)
Efficient errors	0.33	(.20, .44)	0.14	(.00, .28)
Search errors	0.33	(.20, .44)	0.09	(.00, .22)

MZ, monozygotic; DZ, dizygotic; BCS, before a correct sort; ACS, after a correct sort. 95% confidence intervals in parentheses. n = 712

Table 7

ACE Univariate Results-Traditional and Narrowly Defined Errors

Measure	a^2 (CI)	c^2 (CI)	e^2 (CI)	χ^2	AIC
<u>Traditional indices</u>					
Perseverative errors	.35 (.00, .54)	.09 (.00, .39)	.56 (.46, .67)	7.91	-4.09
Nonperseverative errors	.36 (.17, .47)	.00 (.00, .15)	.64 (.53, .76)	6.75	-5.25
Number of trials	.42 (.08, .56)	.04 (.00, .33)	.54 (.45, .65)	1.69	-10.31
<u>Narrowly defined errors</u>					
Perseverative errors BCS	.31 (.00, .43)	.01 (.00, .31)	.68 (.57, .80)	1.42	-10.58
Perseverative errors ACS	.34 (.00, .45)	.01 (.00, .32)	.65 (.55, .78)	0.66	-11.34
Repeated nonperseverative errors	.10 (.00, .37)	.14 (.00, .32)	.76 (.63, .88)	3.61	-8.39
Breaking set errors	.37 (.13, .48)	.00 (.00, .19)	.63 (.52, .75)	7.27	-4.73
Efficient errors	.33 (.03, .44)	.00 (.00, .24)	.67 (.56, .79)	7.69	-4.31
Search errors	.29 (.06, .41)	.00 (.00, .17)	.71 (.59, .84)	10.64	-1.36

a^2 , additive genetic influences; c^2 , shared environment influences; e^2 , nonshared environmental influences which includes error variance; CI, 95% confidence intervals in parentheses; χ^2 , the difference between minus twice the log likelihood (-2LL) of the saturated model and -2LL of the ACE model; AIC, Akaike's Information Criterion. $df = 6$

Table 8

Multivariate Results-Narrowly Defined Errors

Model	Overall Fit				Model Difference Test			
	χ^2	df	AIC	RMSEA	vs.	$\Delta\chi^2$	Δdf	p
A1. Cholesky, ACE	129.07	111	-639.98	0.021				
A2. Cholesky, AE	134.16	132	-676.89	0.007	A1	5.09	21	1.000
B. Two Correlated Genetic Factors, Cholesky E	142.92	139	-682.13	0.009	A2	8.76	7	0.270
C. One Genetic Factor, Cholesky E	145.72	141	-683.33	0.009	B	2.8	2	0.247
D. One Genetic Factor, No Specifics, Cholesky E	153.82	147	-687.23	0.011	C	8.1	6	0.231

χ^2 , the difference between minus twice the log likelihood (-2LL) of the saturated model and -2LL of the reduced model; df, degrees of freedom; AIC, Akaike's Information Criterion; RMSEA, root-mean-square error of approximation; Difference statistics were acquired from comparing models to the model in the versus column. Vs., versus; $\Delta\chi^2$, change in chi-square; Δdf , change in degree-of-freedom; p, p-value; A, additive genetic influences; C, shared environment influences; E, nonshared environmental influences which includes error variance. Best fitting model in bold.